

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 22:26:13 ; Search time 22.75 Seconds  
(without alignments)  
461.936 Million cell updates/sec

Title: US-09-887-038-3  
Perfect score: 2426  
Sequence: 1 MTWQTLTFAHYQPQOMGHS.....WQPPSKQLPPEAEHSDEKM 467

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117.5	4.8	3011	3 US-08-811-566-20	Sequence 20, Appl
2	117.5	4.8	3012	3 US-08-811-566-2	Sequence 2, Appl
3	115.5	4.8	505	1 US-08-221-750A-5	Sequence 5, Appl
4	114.5	4.7	2261	4 US-08-444-818-66	Sequence 66, Appl
5	114.5	4.7	2436	4 US-08-444-818-75	Sequence 75, Appl
6	114.5	4.7	2772	4 US-08-444-818-89	Sequence 89, Appl
7	114.5	4.7	2894	2 US-08-466-975A-23	Sequence 23, Appl
8	114.5	4.7	2894	2 US-08-391-671A-23	Sequence 23, Appl
9	114.5	4.7	2894	3 US-08-467-902A-23	Sequence 23, Appl
10	114.5	4.7	2894	4 US-09-275-265-23	Sequence 23, Appl
11	114.5	4.7	2995	4 US-08-444-818-138	Sequence 138, App
12	114.5	4.7	3011	1 US-08-440-103-36	Sequence 36, Appl
13	114.5	4.7	3011	1 US-08-440-342-36	Sequence 36, Appl
14	114.5	4.7	3011	1 US-07-910-760-10	Sequence 10, Appl
15	114.5	4.7	3011	1 US-08-440-519-10	Sequence 10, Appl
16	114.5	4.7	3011	1 US-08-231-368-36	Sequence 36, Appl
17	114.5	4.7	3011	1 US-08-440-210-36	Sequence 36, Appl
18	114.5	4.7	3011	4 US-09-388-874-2	Sequence 2, Appl
19	114.5	4.7	3011	4 US-09-046-604-36	Sequence 36, Appl
20	113.5	4.7	3011	4 US-09-014-416-1	Sequence 1, Appl
21	110.5	4.6	2955	4 US-08-444-818-124	Sequence 124, App
22	110	4.5	3033	1 US-07-925-695-9	Sequence 9, Appl
23	109	4.5	829	4 US-08-444-818-69	Sequence 69, Appl
24	109	4.5	1786	4 US-08-444-818-54	Sequence 54, Appl
25	107.5	4.4	454	4 US-08-444-818-73	Sequence 73, Appl
26	107.5	4.4	663	3 US-08-824-057-3	Sequence 3, Appl
27	107.5	4.4	3011	1 US-08-453-552-2	Sequence 2, Appl

28	107.5	4.4	3011	2 US-08-710-637-2	Sequence 2, Appl
29	107.5	4.4	3011	5 PCT-US93-00907-2	Sequence 2, Appl
30	106.5	4.4	2955	2 US-08-443-260-3	Sequence 3, Appl
31	106.5	4.4	2955	3 US-08-442-805A-3	Sequence 3, Appl
32	106.5	4.4	2955	3 US-08-443-900A-3	Sequence 3, Appl
33	106.5	4.4	2955	4 US-08-249-843-3	Sequence 3, Appl
34	106.5	4.4	3011	2 US-08-833-678A-6	Sequence 6, Appl
35	106.5	4.4	3011	4 US-08-444-818-17	Sequence 17, App
36	106.5	4.4	3011	4 US-08-529-169A-6	Sequence 6, Appl
37	106.5	4.4	3011	5 US-07-925-695-8	Sequence 8, Appl
38	106	4.4	3033	1 US-09-014-416-3	Sequence 3, Appl
39	105.5	4.3	3010	4 US-09-014-416-5	Sequence 5, Appl
40	104	4.3	3011	4 US-08-188-281B-12	Sequence 12, Appl
41	101.5	4.2	1648	1 PCT-US94-07280-12	Sequence 12, Appl
42	101.5	4.2	1648	5 PCT-US95-01087-12	Sequence 12, Appl
43	101.5	4.2	3011	1 US-08-188-281B-1	Sequence 1, Appl
44	101.5	4.2	3011	1 US-08-453-552-1	Sequence 1, Appl
45	101.5	4.2	3011	1 US-08-453-552-1	Sequence 1, Appl

## ALIGNMENTS

```
RESULT 1
US-08-811-566-20
; Sequence 20, Application US/08811566
; Patent No. 6127116
;
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; NUMBER OF INVENTIONS: 21
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESS: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentrln Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,566
; FILING DATE: 03-MAR-1997
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1113-1-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
;
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
;
; US-08-811-566-20
```

Query Match 4.88; Score 117.5; DB 3; Length 3011;  
Best Local Similarity 20.28; Pred. No. 0.022;  
Matches 100; Conservative 59; Mismatches 158; Indels 177; Gaps 23;  
QY 66 IGLGIATAIAVM---ALLSTIDIDLRQATPIHMYVLLYWGVDALATGILSPVRAALVGLA 122

```

Db 705 VGSIIASMAIAKMEYVILILFLILADARVCCSL-IMMILLISQAIALENLYILMAASLACTH 763
QY 123 KLTLILVFAIAARVLNRPILRSILFSSVYVITSLFVSYGLNQWITGYEEELATWDRNSV 182
Db 764 GLVSELFVECFEA-----WTLKGRVPGA-----786
QY 183 ADFTSRYSYVLGN-PNLLAAYLVPTTAFS-----AAATGVMGMGLPKLLATATACASSLC 236
Db 787 -----YAAEYGMPLILLILLLALPORAAYLDELVAASCC-----GVVLGLMALT 830
QY 237 LILTFYRSGMGLFVAMIFVALLGLFWPDRPLPAPMRRLPFV-VLGGILVAV-----287
Db 831 LSPYTKR-----YISCMWMLQYFLELRVDAOLHVVPPILNVRGGRAVILLMCVYH 881
QY 288 -----LLVAVLG-----LEPLRVKVISIF-----VGRDSSNNFRINWIAV 324
Db 882 PTLVEFITKILLILAFGLPLMIIQASLLKLVKPYFVAVQCLLRICALARKRIAGHY--YQNAI 938
QY 325 LQMIDPRMVLGIGBPNTAFENLVPLVYQOARFRLASVIPLE-----V 367
Db 939 IK-----LGALITGYVYNHPLPLRDMAH-NGLRDLAAVAVPVFSRMEKTLITWGA DT 990
QY 368 AVEGGLL-----GLTAPAVILILVYAVTAVPROVS-----LR 398
Db 991 AAGCDITINGLPVSNARQGOELLILGPADGOMYSKGRILL-APITAYAOOTRGLGCITTSILG 1049
QY 399 RDRN-----PQAFWMLNASLAGLACMLGHLPTVLYRDEASTL-IMWLCIGAI 445
Db 1050 RDKNQVGEQVQIVSTATQTF-LATCINGVCMVYVHGAGRTITASPKGPVQIQTWVNDOL 1108
QY 446 SFW-QPOPSKOLPP 458
Db 1109 VGWPAPOGSRSLTP 1122

```

RESULT 2  
 US-08-811-566-2  
 Sequence 2, Application US/08811566  
 Patent No. 6127116  
 GENERAL INFORMATION:  
 APPLICANT: Rice, Charles et al.  
 TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C  
 TITLE OF INVENTION: VPROS (HCV) AND USES THEREOF  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: David A. Jackson, Esq.  
 STREET: 411 Hackensack Ave, Continental Plaza, 4th  
 STREET: Floor  
 CITY: Hackensack  
 STATE: New Jersey  
 COUNTRY: USA  
 ZIP: 07601  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/811,566  
 FILING DATE: 03-MAR-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jackson Esq., David A.  
 REGISTRATION NUMBER: 26,742  
 REFERENCE/DOCKET NUMBER: 1113-1-006  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-487-5800  
 TELEFAX: 201-343-1684  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3012 amino acids  
 TYPE: amino acid

```

; STRADEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-08-811-566-2

```

Query Match	4.8%;	Score 117.5;	DB 3;	Length 3012;
Best Local Similarity	20.2%;	Pred. No. 0.022;		
Matches 100;	Conservative 59;	Mismatches 158;	Indels 177;	Gaps 23;

[illegible]

```

1      RESULT 3
2      ; Sequence 5, Application US/08221750A
3      ; Patent No. 5643747
4      ;
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Baker, Steven M.
7      ; APPLICANT: Delich, Robert A.
8      ; TITLE OF INVENTION: Games for the Export of Pertussis
9      ; TITLE OF INVENTION: Holotoxin
10     ; NUMBER OF SEQUENCES: 13
11     ;
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
14     ; STREET: Two Millitia Drive
15     ; CITY: Lexington
16     ; STATE: MA
17     ;
18     ; COUNTRY: USA
19     ;
20     ; ZIP: 02173
21     ;
22     ; COMPUTER READABLE FORM:
23     ; MEDIUM TYPE: Floppy disk
24     ; COMPUTER: IBM PC compatible
25     ; OPERATING SYSTEM: PC-DOS/MS-DOS
26     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
27     ;
28     ; CURRENT APPLICATION DATA:
29     ; APPLICATION NUMBER: US/08/221.750A

```

```

; FILING DATE: 31-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/031,619
; FILING DATE: 15-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC93-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-221-750A-5

```

```

Query Match          4.8%; Score 115.5; DB 1; Length 505;
Best Local Similarity 25.7%; Pred. No. 0.0026;
Matches 96; Conservative 44; Mismatches 121; Indels 113; Gaps 21;

```

```

QY 77 WALLSTEDID-----LROATPIHMLVLYWGVDAATGSPVRAALVGLAKLTLY 127
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 WQAAMTSYDAAFSDLAGLRNAPRR---MEATYRLRLQLOPLVYSMTLLVLTGY 135
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 LLYFALAARLR--NPLRLSLFSVVYVITSLFVSV--YGLNOMIYGEELATWDRNSVA 183
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 ----ALLARDRPFEMIRHALL--VAVYVITSLSPDRY--LSTVAVAGVDVAGMLSGPWA 189
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 DFTSRVSYLGNPLAAY--LVPTAFSAAGVWRGMLPKLLAATGASS----- 234
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 -----PDGAGRGGLAQLODQFAQA---QAWVAQLAGQAANDANGSAVMNL 233
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 LC-LILTYSGMIGFVAMIFVWALLGLYWFQRLPAPWRMLFPVGLGVALLVAVYL 293
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 LCAMIVAAAGWGLCLAASL-----LIVRLITLTL---L 265
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 294 GLEPLRVYLSIVGREDSSNFRIVWL-----AVLQMIODRPWLGIGPGNTAFMLVYP 348
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 SLGPIEL--VLLLEPALQRLT---NAMLGALVRLVFPALGTPAVGLLSDVLGALPAG 319
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 349 LYQO-----ARFTLSAYSVLEVAEGGLLGTAFAMLLVTAFAVAVQVSR--LRRDN 402
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 LPRQATEPLRSTMLAA-----TLCATATMLTLVPLASSVNAGLRR--- 362
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 403 PQAFWIMASLAGIA 416
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 --RLMPNAAHPGLA 374
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
; US-08-444-818-66
; Sequence 66, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutler, William J.
; TITLE OF INVENTION: NABV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chilton Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2261 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-66

```

```

Query Match          4.7%; Score 114.5; DB 4; Length 2261;
Best Local Similarity 20.6%; Pred. No. 0.029;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

```

```

QY 66 IGIGLAIAAYW--ALLSLTDILRQATPIHMLVLYWGVDAATGSPVRAALVGLA 122
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 VGSSTISMAIKKEYVYLLLELLADARVCSCL--WMMLISQAEALENLYLAASLAGTH 138
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 KTLVLYVFAALARVLRNPLRLSLFSVVYVITSLFVSVYGLNOMIYGEELATWDRNSV 182
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 GIVSFLVFCFA-----WYLKQKWPGA----- 161
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 ADFTSRVSYLGN--PRLLAAYLVPTTAPS-----AAAGVWRGMLPKLLAATGASSLC 236
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 -----VYTFYGMWPLLLLLLPQRAVALDTEVAASG-----GVLVGLMALT 205
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 237 LILTYSGMIGFVAMIEFWALLGLYWFQRLPAPWRMLFPV--VLGGVAV----- 287
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 LSPYTKR-----YISWCMWMLQYFLTRVEADLHWIPLNVRGGRADVILMCVH 256
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 -----LLVAVLG-----LEPLRVYLSIF---VGRDSSNFRIVMWLAV 324
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 PTLVFDITKLLAVFGPLMIILQASLLKVPYEVKVGGLRFCALARKMIGHY----- 308
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 LQMIODRPWLGIGPGNTAFNLYPLYQARFTALSAYSVLE-----V 367
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 VQWVITK--LQALGTYYVNHLLPLRDMAH--NGLRDLAVAAVEPVFSQMETKLTWGA DT 365
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 368 AVEGGL-----GLTAFAMLLVTAFAVAVQVSR-----LR 398
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 AACGGDILNPLYSARRGRELILGADGVNSKGRLL--APITAYNQOTGSLGCTITSLTG 424
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 399 RDRN-----PQAFWIMASLAGLGHGFLDYVLYRPEASTL--WMLCIGATA 445
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 RDRNVEGEVOIVSTAAQTF--LATCINGCVMTVHGAGCTRTIASPKGFIOMYTNVDDL 483
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 SFN--OPQSKOLPP 458
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 VQWPAQGSRLTP 497
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 5
; US-08-444-818-75
; Sequence 75, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutler, William J.

```

```

1      TITLE OF INVENTION:  NANBV Diagnostics and Vaccines
2      NUMBER OF SEQUENCES:  777
3      CORRESPONDENCE ADDRESS:
4      ADDRESSEE:  Chiron Corporation
5      STREET:  4560 Horton Street
6      CITY:  Emeryville
7      STATE:  CA
8      COUNTRY:  USA
9      ZIP:  94608-2916
10     COMPUTER READABLE FORM:
11     MEDIUM TYPE:  Floppy disk
12     COMPUTER:  IBM PC compatible
13     OPERATING SYSTEM:  PC-DOS/MS-DOS
14     SOFTWARE:  Patent In Release #1.0, Version #1.30
15     CURRENT APPLICATION DATA:
16     APPLICATION NUMBER:  US/08/444,818
17     FILING DATE:
18     CLASSIFICATION:  424
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER:  US/08/403,590
21     FILING DATE:  14-MAR-1995
22     ATTORNEY/AGENT INFORMATION:
23     NAME:  Hardin, Alisha A.
24     REGISTRATION NUMBER:  33,895
25     REFERENCE/DOCKET NUMBER:  0110,002
26     TELECOMMUNICATION INFORMATION:
27     TELEPHONE:  (508)359-3876
28     TELEFAX:  (508)359-3885
29     INFORMATION FOR SEQ ID NO:  75:
30     SEQUENCE CHARACTERISTICS:
31     LENGTH:  2436 amino acids
32     TYPE:  amino acid
33     TOPOLOGY:  linear
34     MOLECULE TYPE:  protein
35     OS-08-444-818-75

```

```

Query Match: 4.7%; Score 114.5; DB 2 Length 2894;
Best Local Similarity 20.6%; Pred. No. 0.042;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

QY 66 LGIGLAAIAAYW---ALSLTDIDIRQAPTHMLVLLWGVDAIATGLSPFAALVGLA 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 705 VGSISASMAIKMEYVLLFLFLADARVCSCL-WMMLLISOEALELVILMAASLATH 763
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 KLTLYLLVFLAARVLRNRLSLFLSVVVTLSLFSVYGINOMIVGVEELAWVDRNSV 182
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 764 GLVSELFVFCFA-----WYLGKKNVFCFA----- 786
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 ADETSRYVSYLGN-PNLLAAYLVPTTAS-----AAAIGWRGWLPRLLAIAATGASSLC 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 787 -----VYTFYEGKMPPLLLLLLPORAALDLTEVAASCG-----GVVLVGLMALT 830
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 237 LILFYSRGGHGLGFVNAIFWVALLGTYWQPRPLPAWRMRLLFPV-VLGGILAV----- 287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 831 LSPYKR-----YISWCLMHLQYLTLYREQLHWMVLPRLNVRGGRDAVYLLMCANH 881
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 -----LVNAVIG-----LEPLRVYLSIF-----VGRDESSNFRINWYLA 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 882 PLTFEDTKLLFLVEGFLMLDASLLKPYEVVRVOGLRRCALARKMIGHY----- 933
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 LQMIODPMLVIGSPGFANLYPLVYQGARFTALSAVSPLP-----V 367
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 934 VQWYIIR--LGALTGYVYNNHLPRLDMAN--NGLRDLAVAVEPVVSOMETKLLTWGADT 990
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 368 AVEGGL-----GLTAFAMLLVYAVTAVAROVS-----LR 398
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 991 AACGDIIINGLPVSARRGREILLGPADGVVSGKGRLL-APITAAQOTRGILGCIITSILTG 1049
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 399 RDRR-----POAFMLASLADGLAMLGHLFDPLVLYRPEASTL-WMLCIGALA 445
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1050 RDKRQVGEVQIVSTAAQVLF-LATCINGVCTVYHGAGTFTIASPKGVIAWITNVDDLL 1108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 SFW-QPQPSKQLPP 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1109 VGWPAPQGRSLTP 1122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
; US-08-391-671A-23
; Sequence 23, Application US/08391671A
; Patent No. 5922532
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLLET, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSWUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHAYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,671A
; FILING DATE: 21-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/920,286

```

```

FILING DATE: 14-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90124241.2
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164100
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-391-671A-23

```

```

QY 66 LGGLIAIAAYW---ALTSLTIDLRQAPTHIMVLTWAGVDLATATGSPRAAAYVGLA 122
Db 705 VGSISMAWKWKYYVLLFLLLADARVCSCL-WMMLLSQEALEENLVIIIMASLAGTH 763
QY 123 KLTLYLLVFLAARVLNRRLRSLLFSVVVITSLFVSYYGLNQMITYGVEELATWDRNV 182
Db 764 GLVSEVLFCEFA-----WYLGKMWPGA-----786
QY 183 ADFLSRYSLTGN-PMLLAALVPTTAFS---AAATGWKRGMLPKLLIAATGASSLC 236
Db 787 -----VYTFYGGWMPLLLLLLALPORAVALDREVAASCG-----GVVLGMLALT 830
QY 237 LILYSRGCGWIGFVAMIFWVALLGTYWFOPLPAPBRMLPPV-VLGGIVAV-----287
Db 831 LSPYTKR-----YISMCMLMDQYFLITRYEADLHWIIFPLNVRGGRDAVILLMCVH 881
QY 288 -----LLVAVLC-----LEPLRVLSIF---VGRDESSNNFRINWLAIV 324
Db 882 PTEVFDITKLLAVFSPMLIDASLLKAYEFYRVGGLRFGCALRKMIGHY-----933
QY 325 LQMIQDQPMWIGIGPMTAFNLVYPLYOQARFALSAVSPLE-----V367
Db 934 VQMYTIK--LGAITGYVYNHLLPLRDMAN--NGLDLDLAVNEPVPVFSOMETKLTWGADT 980
QY 368 AVEGGLL-----GITAFAMLLVYAVTAVROVSR-----LR368
Db 991 AACDDIINGLPVSARGREIILGPADGVMYSKGMRL-APITAYAOQTFRGLLGCIITSLTG 1049
QY 399 RDRN-----PQAFMLASLAGLAGLNGLEDVLYLPRPASTL-WMLCTAIA 445
Db 1050 RDKMVEGEVOJYSTAQT-FLATCINGVCMWTYHGAGTRTIASPKGPVIOMYTNVDOL 1108
QY 446 SFM-QOPSKOLPP 458
Db 1109 VGMFAPOGSKSLTP 1122

RESULT 9
US-08-467-902A-23
: Sequence 23, Application US/08467902A
: Patent No. 6007982
: GENERAL INFORMATION:
: APPLICANT: DELEYS, ROBERT J

```

```

1  APPLICANT: MARLET, DIRK
2  APPLICANT: VAN HEUVESMUN, HUGO
3  TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
4  TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
5  NUMBER OF SEQUENCES: 23
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: NIXON & VANDERHUE, P.C.
8  STREET: 1100 NORTH GLEBE ROAD
9  CITY: ARLINGTON
10 STATE: VA
11 COUNTRY: USA
12 ZIP: 22201
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.30
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/467,902A
22 FILING DATE:
23
24 CLASSIFICATION:
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US/08/391,671
27 FILING DATE:
28 APPLICATION NUMBER: US 07/920,286
29 FILING DATE: 14-OCT-1992
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: WO PCT/EP91/02409
32 FILING DATE: 13-DEC-1991
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: EP 90124241.2
35 FILING DATE: 14-DEC-1990
36 ATTORNEY/AGENT INFORMATION:
37 NAME: SADOFF, B.J.
38 REGISTRATION NUMBER: 36,663
39 REFERENCE/DOCKET NUMBER: 1487-5
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: 7038164000
42 TELEFAX: 7038164100
43 INFORMATION FOR SEQ ID NO: 23:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 2894 amino acids
46 TYPE: amino acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49 MOLECULE TYPE: peptide
50 HYPOTHEICAL: NO
51 ANTI-SENSE: NO
52
53 US-08-467-902A-23
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

```

PRIORITY APPLICATION DATA: EP 90124241.2
APPLICATION NUMBER: 14-DEC-1990
FILING DATE: 14-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2894 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEetical: NO
ANTI-SENSE: NO
US-08-467-902A-23

Query Match 4.7%; Score 114.5; DB 3; Length 2894;
Best Local Similarity 20.6%; Pred. NO. 0.042;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23.

QY 66 LGGLIAIAIAVW---ALLSLTFDIDLRQATPIHMLVLLYWGVDALATGLSPVRAAALVGLA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 705 VGSSIASMAIKWIEVYVLLFLTLADARVCSCLWMMLLSIQEALEMLVILNMAASLAGTH 763

QY 123 KLTIYLLVFLAARVLNRPRLRSLLFSVVYITLSLFVSYYGLNOMIYIVVEELATWVDNRVY 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 764 GLVSLFEFFCFEA-----WYLRGKRWPEA----- 786

QY 183 ADFTSRVYSVLGN-PRILAAVLPPTTAFS-----AAAGVWRGVLPRILAIATGASSIC 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 787 -----VYITYGWMPILLILLALPQRATALDTEVAASG-----GVLYGLMALIT 830

QY 237 LIITYSRGWLGFMVAMIFWVALLGLYWFQPRIPAEWRRLFEVY-VIGGLVAV----- 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 831 LSPYKR-----YISMCLEMLIQLYELTRVEAOIHWMIPEPLNVRGGRDAVILLCAVH 881

QY 288 -----LLVAVLG-----LEPLARVLSIT-----VEREDSSNNPFINWMLAV 324

```

```

Db      882 PTVEDITKLLAVGPIMLDASLLKVPYFVRVGGLLRFLCALKMKMGH----- 933
Qy      325 LQMODRPMGLGPGNTAFNLVPLVQARFALSAYSPLE-----V 367
Db      934 VQWVIIR--LQALGTGVYVHNLPLRDMAH--NGLRDLAVAVEPVVSQMETKLTWGADT 990
Qy      368 AVEGGL-----GLTAFAMLLVTAATAVAVQVSR-----LR 398
Db      991 AACGGIINGLVASARGREILLGPADGMSKGRLL-APITVAAQQTGILGCIITSLTG 1049
Qy      399 RDRN-----PQAFMLMASLAGIAGMLGHGFDTVLVYRPEASTL--WMLCIGAIA 445
Db      1050 RDKNQVEGEVOIVSTAQTF--LATCINGVCWTVYHGAGTFTIASPKGPVIOMYTNVDL 1108
Qy      446 SFW-QOPSKOLPP 458
Db      1109 VGMPAPQGSRLTP 1122

RESULT 10
US-09-275-265-23
; Sequence 23, Application US/09275265
; Patent No. 6287761
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT J
; APPLICANT: POLETT, DIRK
; APPLICANT: MAERTENS, GEERT
; APPLICANT: VAN HEUVERSMUN, HUGO
; TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
; TITLE OF INVENTION: ANTIBODIES TO HEPATITIS C VIRUS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEE ROAD
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,265
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,671
; FILING DATE: 21-FEB-1995
; APPLICATION NUMBER: US 07/920,286
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP91/02409
; FILING DATE: 13-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90124241.2
; FILING DATE: 14-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: SADEFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2894 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```

```

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-275-265-23

Query Match          4 7%: Score 114.5; DB 4; Length 2894;
Best Local Similarity 20.6%; Pred. No. 0.042;
Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

Qy      66 LGIGIAIAAW---ALLSLTDIDLROATPIHMLVLTYMGVDALATGSPVRAAALVGLA 122
Db      705 VGSSIASMAIKWEVVLPLFLDLARVCSCL--WMMLISQAEALAEVLVILNMAASLAGTH 763
Qy      123 KLTLYLVFALAARVLAENPRLSLFSVYVITSLSFVSYGILNOMIYGEELATWDRNSV 182
Db      764 GLVSFLVFCEFA-----WYLDGKWVPGA----- 786
Qy      183 ADFTSRVSYLGN--PNLLAAYLVPTAFS-----AAAGVWRGMPLPKLLAIAATGASSIC 236
Db      787 -----VTFYGMWPLILLILLALPQRAYALDTEVMAISCG-----GVLVGLMALTP 830
Qy      237 LILTYSRGMLGFVAMLEFVWALLGLYWPQRLPAWRMLFPV--VLGGLVAV----- 287
Db      831 LSPYKR-----YISCMIMWLYEFLTRVEAQHLWVITPLNVGGRDAVILLMCAYH 881
Qy      288 -----LLVAVLG-----LEPLRVYLSIF-----VGRDSSNFRINWMLAV 324
Db      882 PTVEDITKLLAVFGPLWILDSLLKVPYFVRVGGLLRFLCALKMKMGH----- 933
Qy      325 LQMODRPMGLGPGNTAFNLVPLVQARFALSAYSPLE-----V 367
Db      934 VQWVIIR--LQALGTGVYVHNLPLRDMAH--NGLRDLAVAVEPVVSQMETKLTWGADT 990
Qy      368 AVEGGL-----GLTAFAMLLVTAATAVAVQVSR-----LR 398
Db      991 AACGGIINGLVASARGREILLGPADGMSKGRLL-APITVAAQQTGILGCIITSLTG 1049
Qy      399 RDRN-----PQAFMLMASLAGIAGMLGHGFDTVLVYRPEASTL--WMLCIGAIA 445
Db      1050 RDKNQVEGEVOIVSTAQTF--LATCINGVCWTVYHGAGTFTIASPKGPVIOMYTNVDL 1108
Qy      446 SFW-QOPSKOLPP 458
Db      1109 VGMPAPQGSRLTP 1122

RESULT 11
US-08-444-818-138
; Sequence 138, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chilton Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590

```





RESULT 13  
 US-08-440-542-36  
 ; Sequence 36, Application US/08440542  
 ; Patent No. 5670153  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weiner, Amy J.  
 ; APPLICANT: Houghton, Michael  
 ; TITLE OF INVENTION: Immunoactive Polypeptide Compositions  
 ; NUMBER OF SEQUENCES: 45  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Chiron Corporation  
 ; STREET: 4560 Horton Street  
 ; CITY: Emeryville  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94608  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/440,542  
 ; FILING DATE: 12-MAY-1995  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/231,368  
 ; FILING DATE:  
 ; APPLICATION NUMBER: US 07/759,575  
 ; FILING DATE: 13-SEP-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: McClung, Barbara G.  
 ; REGISTRATION NUMBER: 33,113  
 ; REFERENCE/DOCKET NUMBER: 0205,001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (510) 601-2708  
 ; TELEFAX: (510) 655-3542  
 ; INFORMATION FOR SEQ ID NO: 36:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3011 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-440-542-36

Query Match 4.7%; Score 114.5; DB 1; Length 3011;  
 Best Local Similarity 20.6%; Pred. No. 0.045;  
 Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

QY	66	LGIGIAIAAYW---ALSLTIDLRQATPIHVLVLYWGVDALATGSLSPVRAALVGLA	122
DB	705	VGSSTIASMAIKWEYVLLFLLDADAVCSCL-WMMLLISQAEALLENVILMAASLAGTH	763
QY	123	KLTLILYFALAAVLRNRLRSLSFSVYVITSLFVSYGLNQMIYGEELATWDRNSV	182
DB	764	GLVSFLVFFCPA-----WYLGKRWVPGA-----	786
QY	183	ADFTSRVYSYIGN-PNLAAVLYPTAFS-----AAAGVWRGWLPKLIAIAATGASSIC	236
DB	787	-----VYTFGMPPLDLLLALLPQRAVALDTEVAASCG-----GVVLGMLALT	830
QY	237	LILYSRSGWGLFVAMITVMALIGYQFRLPAPWRKMLFPV-VLGLIVAV-----	287
DB	831	LSPIYKR-----YISWCLMWLQFLRYEAOJLHWIIPLANRGGRDAVILLMCAVH	881
QY	288	-----LIVAVLG-----LEPLRYRVLSIF-----VGREDSNNFRIMWMLAV	324
DB	882	PLVVEDIRKLLAVFGPLMIIDASLLKPYFVRVQGLRFPALAKMKMGHT-----	933
QY	325	LQMIODRPMWLGIGPGNTAFNLVPLLYQOARFTALSAYSVPLE-----V	367

DB 934 VQMWIIR--LGALTGTYVYVNHLLPLRDMNH--NGLRDLAAVEPVEFSQMETRLITWGADT 990  
 QY 368 AVEGGIL-----GLTAFAMLLVTAIVAVQVSR-----LR 398  
 DB 991 AACGDITNGLPVSARGREILGPADGWVSKGMRLL-APITAYAOQTRGLLGIITISLTG 1049  
 QY 399 RDRN-----POAFWIMASLAGMIGHGLDVTLYRREASTL-WMLCIGIAIA 445  
 DB 1050 RDKNOVEGEVQIVSTRAOTF-LATCINGCWTYVHAGAGTRTASPGVIOGMTVNDODL 1108  
 QY 446 SFW-QPQPSKOLPP 458  
 DB 1109 VGMPAPQGSRSRLTP 1122

RESULT 14  
 US-07-910-760-10  
 ; Sequence 10, Application US/07910760  
 ; Patent No. 5683864  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Houghton, Michael  
 ; APPLICANT: Choo, Qui-Lim  
 ; APPLICANT: Kuo, George  
 ; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
 ; TITLE OF INVENTION: Antigens for Use in Immunoassays for Anti-HCV Antibodies  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Chiron Corporation  
 ; STREET: P.O. Box 8097 (Int. Prop. R-440)  
 ; CITY: Emeryville  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 94662-8097  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/910,760  
 ; FILING DATE: 07-JUL-1992  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Blackburn Esq., Robert P.  
 ; REGISTRATION NUMBER: 30,447  
 ; REFERENCE/DOCKET NUMBER: 0101,002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (510) 601-2702  
 ; TELEFAX: (510) 655-3542  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3011 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-07-910-760-10

Query Match 4.7%; Score 114.5; DB 1; Length 3011;  
 Best Local Similarity 20.6%; Pred. No. 0.045;  
 Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;

QY	66	LGIGIAIAAYW---ALSLTIDLRQATPIHVLVLYWGVDALATGSLSPVRAALVGLA	122
DB	705	VGSSTIASMAIKWEYVLLFLLDADAVCSCL-WMMLLISQAEALLENVILMAASLAGTH	763
QY	123	KLTLILYFALAAVLRNRLRSLSFSVYVITSLFVSYGLNQMIYGEELATWDRNSV	182
DB	764	GLVSFLVFFCPA-----WYLGKRWVPGA-----	786
QY	183	ADFTSRVYSYIGN-PNLAAVLYPTAFS-----AAAGVWRGWLPKLIAIAATGASSIC	236
DB	787	-----VYTFGMPPLDLLLALLPQRAVALDTEVAASCG-----GVVLGMLALT	830

QY 237 LILFYSRGWLGFWAMIFVWMLLGLVWFOPLPAPWRRLFPV-VLGSLVAV----- 287  
 DB 831 LSPYKR-----YISWCLMWLOFLTRVEAOHLHWITPLVHGRDAVITLLMCVH 881  
 QY 288 -----LLVAVLG-----LEPLRVVLSIF---VGRDSSNNFRINWMLAV 324  
 DB 882 PTLVFDITKLLAVFGPLMILQASLLKVPYFVRVQGLRFCALARKMIGHY----- 933  
 QY 325 LQMIODRPMUGIGPGNTAFNLYPLIQOARFTALSAYSPLF-----V 367  
 DB 934 VQWVYIK--LGAALTGYVYVNHLPRLDMWH-NGLRDLAAVEPVSQMETKLIITWGADT 990  
 QY 368 AVEGGL-----GLTAFAMLLVTAVTAVROVSR-----LR 398  
 DB 991 AACGDIINGLVYSARRGREILLGPADGMSKGRLL-APTITAAQOTRGLGCIITSLTG 1049  
 QY 399 RDRN-----POAFWMLASLAGMLGHGLFDVLYRPEASTL-WMLCIGAIA 445  
 DB 1050 RDKNQEVEQIVSTAQTF-LATCINGVCWTVYHAGTRTASPKGPVIQMTNVDDOL 1108  
 QY 446 SFW-QPOPSKQLPP 458  
 DB 1109 VGWPAPOGSRSLTP 1122

RESULT 15  
 US-08-440-519-10  
 ; Sequence 10, Application US/08440519  
 ; Patent No. 5712087  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Houghton, Michael  
 ; APPLICANT: Choo, Qui-Lim  
 ; APPLICANT: Kuo, George  
 ; TITLE OF INVENTION: Combinations of Hepatitis C virus (HCV)  
 ; NUMBER OF INVENTION: Antigens for use in immunoassays for Anti-HCV Antibodies  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Chiron Corporation  
 ; STREET: P.O. Box 8097 (Int. Prop. R-440)  
 ; CITY: Emeryville  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 94662-8097  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/440,519  
 ; FILING DATE: 12-MAY-1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/910,760  
 ; FILING DATE: 07-JUL-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Blackburn Esq., Robert P.  
 ; REGISTRATION NUMBER: 30,447  
 ; REFERENCE/DOCKET NUMBER: 0101.002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (510) 601-2702  
 ; TELEFAX: (510) 655-3542  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3011 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-440-519-10

Query Match 4.7%; Score 114.5; DB 1; Length 3011;

Best Local Similarity 20.6%; Pred. No. 0.045;  
 Matches 102; Conservative 55; Mismatches 160; Indels 177; Gaps 23;  
 QY 66 LGIGLAIATAYW---ALSTITDIDLRQATPIHMLVLLYVWVDLALGSLPVRALAYGLA 122  
 DB 705 VGSISASWALKMEYVYLLFFLLADARVCSCL-WMMLLSQAEALENLVITLNASLGGTH 763  
 QY 123 KLFLLVLFALAAVRLNPNRLSLFSSVYVTSLSFVSYGLOMIVGVELATWVDNSV 182  
 DB 764 GLVSLVFCEFA-----WYIKGWVGA----- 786  
 QY 183 ADFTSRVSYLGN-PULLAAYVPTAFS-----AAAGYWGMLPKLLAIAATGASSIC 236  
 DB 787 -----VYTFYGMWPLILLALLPQRAYALDTEVAASCG-----GVVLVGLMALT 830  
 QY 237 LILFYSRGWLGFWAMIFVWMLLGLVWFOPLPAPWRRLFPV-VLGSLVAV----- 287  
 DB 831 LSPYKR-----YISWCLMWLOFLTRVEAOHLHWITPLVHGRDAVITLLMCVH 881  
 QY 288 -----LLVAVLG-----LEPLRVVLSIF---VGRDSSNNFRINWMLAV 324  
 DB 882 PTLVFDITKLLAVFGPLMILQASLLKVPYFVRVQGLRFCALARKMIGHY----- 933  
 QY 325 LQMIODRPMUGIGPGNTAFNLYPLIQOARFTALSAYSPLF-----V 367  
 DB 934 VQWVYIK--LGAALTGYVYVNHLPRLDMWH-NGLRDLAAVEPVSQMETKLIITWGADT 990  
 QY 368 AVEGGL-----GLTAFAMLLVTAVTAVROVSR-----LR 398  
 DB 991 AACGDIINGLVYSARRGREILLGPADGMSKGRLL-APTITAAQOTRGLGCIITSLTG 1049  
 QY 399 RDRN-----POAFWMLASLAGMLGHGLFDVLYRPEASTL-WMLCIGAIA 445  
 DB 1050 RDKNQEVEQIVSTAQTF-LATCINGVCWTVYHAGTRTASPKGPVIQMTNVDDOL 1108  
 QY 446 SFW-QPOPSKQLPP 458  
 DB 1109 VGWPAPOGSRSLTP 1122

Search completed: January 10, 2002, 23:04:50  
 Job time: 2317 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 22:25:53 ; Search time 38.09 seconds  
(without alignments)  
908.171 Million cell updates/sec

Title: US-09-887-038-3  
Perfect score: 2426  
Sequence: 1 MTVMQTLFAHYQPOQMGHS.....WQPPSKQLPEAEHSDEKX 467

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq..1101:.\*  
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	6.0	388	20	AAV22569
2	143.5	5.9	348	21	AAV81520
3	132.5	5.5	758	22	AAV79584
4	132.5	5.5	962	22	AAV90046
5	130	5.4	536	21	AAV4555
6	127	5.2	461	18	AAV19276
7	124.5	5.1	483	22	AAV89834
8	124.5	5.1	483	22	AAV78975
9	124.5	5.1	553	20	AAV38796
10	123.5	5.1	604	21	AAV75522
11	122.5	5.0	558	20	AAV38797

12	121	5.0	386	21	AAV90863
13	119.5	4.9	604	21	AAV75523
14	119	4.9	806	22	AAV92136
15	118.5	4.9	535	21	AAV43136
16	118.5	4.9	535	21	AAV71061
17	118.5	4.9	535	22	AAV72395
18	118	4.9	560	22	AAV81108
19	117.5	4.8	3011	19	AAV77397
20	117.5	4.8	3011	19	AAV77398
21	117	4.8	513	20	AAV38786
22	116	4.8	533	22	AAV72393
23	115.5	4.8	505	18	AAV23275
24	115.5	4.8	604	21	AAV75521
25	115	4.7	3033	14	AAV33214
26	114.5	4.7	2261	10	AAV90164
27	114.5	4.7	2301	10	AAV92047
28	114.5	4.7	2435	13	AAV25135
29	114.5	4.7	2436	10	AAV92050
30	114.5	4.7	2436	10	AAV92050
31	114.5	4.7	2436	13	AAV28582
32	114.5	4.7	2772	11	AAV08123
33	114.5	4.7	2772	21	AAV18540
34	114.5	4.7	2894	13	AAV24440
35	114.5	4.7	2894	16	AAV70230
36	114.5	4.7	2955	20	AAV14975
37	114.5	4.7	2955	21	AAV18541
38	114.5	4.7	3011	13	AAV21519
39	114.5	4.7	3011	14	AAV31621
40	114.5	4.7	3011	17	AAV09031
41	114.5	4.7	3011	18	AAV34480
42	114.5	4.7	3011	19	AAV40038
43	113.5	4.7	2984	22	AAE00442
44	113.5	4.7	2984	22	AAE00447
45	113.5	4.7	2984	22	AAE00449

#### ALIGNMENTS

RESULT 1  
ID AAV22569 standard; Protein: 388 AA.  
XX  
AC AAV22569;  
XX  
XX 17-NOV-1999 (first entry)  
XX  
XX Bacterial general essential protein gep1713.  
XX  
XX Bacterial general essential protein; pathogenic bacteria; pathogen; inhibitor;  
XX  
XX Bacterial growth.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX WO9933871-A2.  
XX  
XX 08-JUL-1999.  
XX  
XX 30-DEC-1998; 98WO-US27918.  
XX  
XX 31-DEC-1997; 97US-0070116.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Youngman P, Fritz C, Murphy C, Guzman L;  
XX  
XX WPI; 1999-430230/36.  
XX  
XX N-PSDB; AA220360.  
XX  
XX Streptococcus pneumoniae general essential protein genes and proteins,  
XX  
XX useful for identification of antibacterial agents -  
XX  
XX Claim 1; Fig 13; 124pp; English.



DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:684.  
 XX  
 XX Corynebacterium glutamicum; carbon metabolism and energy production;  
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KW fine chemical production; organic acid; proteinoenic amino acid;  
 KW nonproteinoenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100844-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000MO-IB00943.  
 XX  
 PR 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031412.  
 PR 08-JUL-1999; 99DE-1031413.  
 PR 08-JUL-1999; 99DE-1031419.  
 PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031431.  
 PR 08-JUL-1999; 99DE-1031433.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031562.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 09-JUL-1999; 99DE-1032180.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 09-JUL-1999; 99US-0143308.  
 PR 14-JUL-1999; 99DE-1032924.  
 PR 14-JUL-1999; 99DE-1032973.  
 PR 27-AUG-1999; 99DE-1033005.  
 PR 31-AUG-1999; 99DE-1040765.  
 PR 03-SEP-1999; 99US-0151572.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042123.  
 PR 03-SEP-1999; 99DE-1042125.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habermayer G;  
 DR N-PSDB: AAF71701.  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 PT metabolism and oxidative phosphorylation protein for production or  
 PT modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes -  
 XX  
 PS Claim 20; Page 1109-1111; 1246pp; English.  
 XX  
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in  
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and  
 CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (II) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinoenic  
 CC or nonproteinoenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins

CC (III) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).  
 XX  
 XX Sequence 758 AA:  
 XX  
 Query Match 5.5%; Score 132.5; DB 22; Length 758;  
 Best Local Similarity 22.5%; Pred. No. 0.0001;  
 Matches 116; Conservative 67; Mismatches 188; Indels 145; Gaps 24;  
 QY 20 SPSFLAR-----LFGSLRAMRASSQLVWSSEALGFLAVVYGSAPFPSSALGIGLAA 72  
 DB 38 satflhaavvkagilyllrlfsivfhdvaywn-----wllily-----gmgtcal 80  
 QY 73 IAAVWALSLTDIDLRQATPHIMVLLVWGVDAIATGSLSPVRAALV-----GLAKTIV 127  
 DB 81 msayfa---vqktdlkkklaystsvshlgwlvatlgvcpfaigaalvhlshalfksslf 137  
 QY 128 LVFVALARVLRNPRSLFSVWVTSLFSVSVYGLNOMIVGEELATWVDNSVADFTS 187  
 DB 138 mligvindhgt-qttrdirrigflvkkmpftfvsv-----lgaismas-----vp 180  
 QY 188 RYVSYIGNPNLLAAVLPPTAFSAAGVWRCMLPRLALAAVWGSLLCLITVSRGWL 247  
 DB 181 pllgfvskegmI-----tafmddpIg--nsyvvlllvgaalge-----vltfysakl 226  
 QY 248 GRVAMIFVWALLGLVWFQPR-----LPAPWRKWLFPVVLGGLVANVLAVALGLEPLRRV 302  
 DB 227 -----vgaftvdgprdmshvkeapvsIwl--paalpgimsiplylvslftapasa 275  
 QY 303 LSLFVREDSSNN-----FRINWLAVLQMTQDRPWLG 336  
 DB 276 aatsaageaaahmlhahwngintpellsigvavagllgvlfrkelw-----klaetspf pi 330  
 QY 337 GCGNTAFN-LVYPLYQOAREFALSAYSVLE---VADEGLGLTFAFA----- 380  
 DB 331 aqndllsmlyranllqkffgmadsmsprhlsvslilwalaafatilpsvqlapkq 390  
 QY 381 -----WLLVTVAVTAVRQVSRRLRDRNPOAFWLMASLAGLACMLGHGL-FDTVLY-REDA 433  
 DB 391 pyidfwldlplajalalsvfgllltnr-----rlsaavlvgtvgyvsfgmlilgabdV 444  
 QY 434 STLAMWLCIGALIASF-----WQPSKOLPPEAHS 463  
 DB 445 altqflvegllvvlmmvrvhnpafkrlkpsrrts 480  
 RESULT 4  
 AAG90046  
 ID AAG90046 standard; Protein; 962 AA.  
 XX  
 AC AAG90046;  
 XX  
 XX 26-SEP-2001 (first entry)  
 DE  
 XX C glutamicum protein fragment SEQ ID NO: 3800.  
 XX  
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EPI108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.

```

XX 16-DEC-1999: 99JP-0377484.
PR 07-APR-2000: 2000JP-0159162.
PR 03-AUG-2000: 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOOGYO KK.
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
DR N-PSDB; AAB65265.
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 3800; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 962 AA;
SQ
Query Match 5.5%; Score 132.5; DB 22; Length 962;
Best Local Similarity 22.5%; Pred. No. 0.00014;
Matches 116; Conservative 67; Mismatches 188; Indels 145; Gaps 24;
OY 20 SSFLHR-----LFGSRANRASSQLLVWSEALGFLNAVYGSAPVPSALGLGLAA 72
DB 242 saflhaavvkvaglylllrlfsvfhdvavvn-----wllilv-----gmgrai 284
OY 73 IAAVWALSLINDIDROATPIHMLVLLWGVDAATATGSPRAAALV-----GLAKLTLY 127
DB 285 msayfa---vqktldkltaystshgwtvavltgvtprfalgaaivthlshalfksl 341
OY 128 LLVFALAARVLRNPRLSLFSVVVTSLEWVSYGLNOMIVGEELATWDRNSVADPTS 187
DB 342 mllygidhgt-gtrdirrlgflvkkmpftfsv-----lglalsmas-----vp 384
OY 188 RYVSYLGNPNLLAATLVTTAFSAATIGVWGMPLKLLAATGSSCLITLTSRSGWL 247
DB 385 pllygfskegm-----tafmadpiz--nsyvvlllvgaalga-----vltfysaki 430
OY 248 GFVAATIFWALIGLWFOPR-----LPAPMRRLFPVVLGGLVAVLLVAVGLEPLFRVY 302
DB 431 -----vlgatvdgprdmshvkeapvskwl-paalpglmslplvlsifdapsa 479
OY 303 LSIWGEDSSNN-----FRINWLAVALQMTQDRPWIGI 336
DB 480 aatsaageaahmlhalwhgintgllisigvavagllgvlfkkelw-----kiaetspf 534
OY 337 GGGNATFN-LVYPLXQOARFTALSAVPLE-----VAVEGULGILTAFA----- 380
DB 535 atgndilsmlyvranllkffigmdmsprthlvslivllwalaafatclipsvapkq 594
OY 381 -----MLLVTAVTAVRQVSRRLRRDRNPOAEFWIMASLAGLMGHG--FDTVLV-REEA 433
DB 595 poidtwidilpialialavsfllttrn-----rlsaavlygtvgvgsfmglllgpav 648

```

```

OY 434 STLWMLCIGAIASF-----WOPQPSKOLPEAEHS 463
DB 649 alqqlveglvvlmmvvrhqpantfklkpsrrrs 684
XX
XX RESULT 5
XX AAB44555
ID AAB44555 standard; Protein; 536 AA.
XX
XX AAB44555;
XX
XX 08-FEB-2001 (first entry)
XX
XX Virulence gene protein #35.
XX
XX Virulence gene; antibacterial; vaccine; bacterial infection;
XX septicemia; bronchopneumonia; rhinitis; wound infection.
XX
XX Pasteurella multocida.
XX
XX MO200061724-A2.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000MO-US09218.
XX
XX 09-APR-1999; 99US-0128689.
XX 10-SEP-1999; 99US-0153453.
XX
XX (PHAA ) PHARMACIA & UPOJHN INC.
XX
XX Lowery DE, Fuller TE, Kennedy MJ;
XX WPI: 2000-647422/62.
XX N-PSDB; AAC79616.
XX
XX Attenuated Pasteurellaceae bacteria comprising mutations in virulence
XX genes, useful as a live attenuated vaccine against bacterial infections
XX
XX Claim 39; Pages 208-209; 322pp; English.
XX
XX The family Pasteurellaceae encompasses several pathogens that infect a
XX wide variety of animals. The present invention relates to virulence genes
XX from Pasteurellaceae. The present sequence is a protein encoded by one
XX such virulence gene. The virulence genes of the present invention may be
XX mutated in order to produce an inactive gene. The inactive virulence gene
XX may in turn be used to produce a vaccine, which is useful for treating
XX bacterial infections such as septicemias, bronchopneumonias, rhinitis and
XX wound infections.
XX
XX Sequence 536 AA;
SQ
Query Match 5.4%; Score 130; DB 21; Length 536;
Best Local Similarity 21.8%; Pred. No. 0.00011;
Matches 95; Conservative 47; Mismatches 139; Indels 154; Gaps 22;
OY 26 LFGSLRA--WRASSQL-----LWSEALGFLNAVYGSAPVPSALGLAIAAYWA 78
DB 1 mfkrtfityrpsaylgmivivfisaifyafalgavf-slpfars-----wt 46
OY 79 LLSLTIDIDROATPIHMLVLLWGVDAATATGLSPVRAALVGLAKLTLYLVFALAARV 138
DB 47 al-lsdqylq-----hvlifsfw-----gaflsavlavllfglivaatafyqpivgkll 95
OY 139 RNPRLSLFSVVVTSLEF--VSYYGLNOMIVGEELATWDRNSVADPTSRYVSYLGNP 196
DB 96 ---klisltivlpalvaligllvgvgasgwlamsigftaw-----dwcprniygltg-- 143
OY 197 NLAAAYL-----VPTTASAAATIGVWGM-----LPKL 224
DB 144 -lllehnlffnvplacrllfglqglpvgqrglaeqnllrgwhffrllewpylrgqllpaf 202

```

```

QY 225 LAIAATGASSLCLITYSRG-----GNLCFVAMIFWALIGL 261
DB 203 tllfmictsfafvltlgygpyktllevalygaalfefdvkagllfallqfvcflflfcl 262
QY 262 YMF-----QRLPAPMRRLFPVVLGGLVANVLAVLGLBPLRVRSIFVGRE 310
DB 263 ssfispapattlhpqfwfapqsyw---vklwgmilvcavtfillpplntlvtsall--- 316
QY 311 DSSNRRINWLAFLQMIQDRPMLGICPGMTARNLVYPLYQOARFTALSAVSPLEVAYE 370
DB 317 --ssgf-ftlwtl-----qpqlwkalg-----ys--ltlapl 342
QY 371 GGLIGLT-AFAMPLL 384
DB 343 sallalvtsfallll 357

RESULT 6
AAM19276
ID AAM19276 standard; Protein; 461 AA.
XX
AC AAM19276;
XX
DT 30-DEC-1997 (first entry)
XX
DE Pseudomonas aeruginosa pilO polypeptide.
XX
KM pilO, pilIn; glycosylation; Gram-negative bacterium; infection;
XX diagnosis; vaccine; O antigen.
XX
OS Pseudomonas aeruginosa strain 1244.
XX
FH Key Location/Qualifiers
FT Misc-difference 64 /note="encoded by CTG"
XX
PN -WO9723600-A1.
XX
PD 03-JUL-1997.
XX
PF 19-DEC-1996; 96WO-US19747.
XX
PR 18-DEC-1996; 96US-0768176.
PR 22-DEC-1995; 95US-0009190.
XX
PA (US9A ) USARMRC US ARMY MEDICAL RES MATERIAL COMMAND.
XX
PI Castirc P, Cross A, Sadoff J;
XX
DR WPI; 1997-351043/32.
DR N-PSDB; AAT72332.
XX
PT DNA encoding Pseudomonas aeruginosa pilO - which glycosylates pilIn,
PT useful to develop products to diagnose and in vaccines for
PT Gram-negative bacterial infections
XX
PS Claim 4; Page 40-42; 60pp; English.
XX

```

```

CC Large-scale production of GNB pill are claimed.
XX
SQ Sequence 461 AA;
XX
Query Match 5.2%; Score 127; DB 18; Length 461;
Best Local Similarity 23.8%; Pred. No. 0.00019;
Matches 98; Conservative 53; Mismatches 152; Indels 108; Gaps 24;
QY 96 LVLLYMGVDALATGLSPVRAALVGLAKLTLYL-----LVFA-----LAARVLRN 140
DB 56 llllfstksfarrmvgvllvllglsvsafslsnpswafkxsvfglmfslfnias 115
QY 141 PR--LRSLFVSVVITSLSVSYGGLNQWITGVBEIATWDRNSVADFTSRV-----YSTLG 194
DB 116 pwwrrlalgvvvlggffcyqgflly-----laafv--sglrelprvllsgfshvr 166
QY 195 NPMLAAVLPPTTAFSAALIGVW-----RGWLPKLLAIAATGASSLCLITYSRGW 246
DB 167 tngqfagmlpl-----maalglylretgfrflswlvmll-laiqwcistala---grqlw 218
QY 247 LCFVAMIFWALLGLYWFQRLPAPMRRLFPVVLGGLVA-----VLVAV--LGLBPL 298
DB 219 lgfav-----ahlaclwlgp-----vgrrflivqlsaafvqalylflmwalptwlgd-- 267
QY 299 RVRLVLSIFVGRDSSNNRRINWLAFLQMIQDRPMLGICPGMTARNLVYPLYQOARFTAL 358
DB 268 ---mclmsgm--sglsitrdlwtadwgmfvahpdllygv-----mhfsa-- 307
QY 359 SAVSPLEVAVEGGLGLTFAFA-W-----LLLVTAVT--AVRQVRLRRDRMPQ--AFML 408
DB 308 ----vpnsygaahpghmllqwfewggaagllvvqimtlgllirgaryllegqpmdaglwl 363
QY 409 MASLAGLGMGLGLEDVLYLRPEASTLMWLCIGAIASFMQ----POPSKO 455
DB 364 -----alvsvlvlaqvdygfvmprfcqvtvllalvglamrwsprvvpagqr 409

RESULT 7
AAG89834
ID AAG89834 standard; Protein; 483 AA.
XX
AC AAG89834;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 3588.
XX
KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127668.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH65053.
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing

```





Db	12	aligvlavltvtwladgslngpdmadlqmsdgumniaavaallfoglft-----	62
Oy	171	EBELAWDRNSVADFTSRV-YSYLGN-PNLAAVLYPTTAFSAAGVWGMLEPKLLAIA	228
Db	63	-----lvaaggladvfgvrvimmignlilnlgslilietatcslatqmwlgvrlqilaad	117
Oy	229	ATGASSLCILLYT-----SNGGW-IGFVAMT-----FWV-----AL	256
Db	118	almssaalavtklywlgtdqravslwslgswgtgcalfaglvvaapfgwrglffalcal	177
Oy	259	LGILYFQRLPALPAMWRMLFPPVYLGL-----VAVLLVAVLGLF-----	296
Db	178	vsivaialtrhipear-----paqslmhlwsgllvialslvelflitqgeslqwthmt	234
Oy	297	----PLRVRVLSIFVGRDSSNNFRIINW-----LA	323
Db	235	wlllavslflavrfllf-----flaewpdlfnlfkdhafsqatlnflmetsgvyva	288
Oy	324	VLQMTQDRPWLIGT-----GNTAFNLVPLYQQAQRTALSAYSVPLEVAEGGLG	375
Db	289	vvmwvqmgw-gvspcltsgltslqfaafvllfivgekamqkvgara-----vlltagllv	343
Oy	376	LTAFAFMILLYAVTVAROVSRRLRDRNRNQATWMLASLAG--LACMLCHGLEFDVLYLRPEA	433
Db	344	alatal-limltafv-----estcyllvslaglsfslg-19lglflapvtdtal	387
Oy	434	STL 436	
Db	388	gfl 390	
RESULT	9		
AAV38796			
ID	AAV38796	standard; Protein: 553 AA.	
XX	AAV38796;		
XX	- 08-OCT-1999	(first entry)	
XX	Neisseria meningitidis antigen encoded by ORF141.		
XX	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;		
KM	treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.		
XX	Neisseria meningitidis.		
OS			
XX	W09924578-A2.		
PN			
PD	20-MAY-1999.		
XX			
XX	09-OCT-1998;	98WO-IB01665.	
PF			
PR	01-SEP-1998;	98GB-0019016.	
PR	06-NOV-1997;	97GB-0023516.	
PR	14-NOV-1997;	97GB-0024190.	
PR	18-NOV-1997;	97GB-0024386.	
PR	27-NOV-1997;	97GB-0025158.	
PR	10-DEC-1997;	97GB-0026147.	
PR	14-JAN-1998;	98GB-0000759.	
FA			
XX	(CHIR-) CHIRON SPA.		
PI	Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;		
XX	WPI: 1999-327407/27.		
DR	N-PSDB; AA212229.		
XX			
XX	Proteins from Neisseria meningitidis and N. gonorrhoeae useful for		
PT	diagnosis, treatment and prevention of infection		
XX			
PS	Claim 4; Page 338-339; 52App; English.		
XX			
CC	Amino acid sequences AAV38499-Y38944 represent Neisseria meningitidis		

CC and N gonorrhoeae antigenic proteins. They are encoded by open CC reading frames (ORFs) A241972-21258. The antigenic proteins, CC their fragments, their nucleic acids and antibodies are used for CC diagnosis, prevention (as vaccines) or treatment of Neisseria CC infections, such as meningitis, septicaemia and gonorrhoea. Both CC organisms are closely related. Fragments of the nucleic acids CC are useful as hybridisation probes and antisense reagents.

**SQ Sequence 553 AA;**

Query Match	5.18;	Score 124.5;	DB 20;	Length 553;
Best Local Similarity	24.88;	Pred. No. 0.00043;		
Matches 109;	Conservative 43;	Mismatches 141;	Indels 147;	Gaps 23;

OY	79	ILSLFDIDLROATPTIH-----MLVLL-----	YMGVDALATGLSP	112			
		: :	: :				
Db	1	mltlypparpraktchekpylllmafaawlpvgvfhshlwnpdeparavytveaalagsip		60			
OY	113	VRAAALVGLAKL-----TLVTLVFPALAAVRLRNP-----	RLSLFSVVITSLFVS	160			
		: :	: :				
Db	61	l-vahlfgqldfglpppyllw-aaafkhlispwaadsydaarfagvffavlgltscgfag		118			
OY	161	YGLNOMITGVEBELTWDNRNSVADFTSRVSYGCPMLL--AAVLPPTAFSAALGWR		218			
		: :	: :				
Db	119	fnflgrhbg-----rsvvlilqclglfpvahflmpaaaafaaagvlh		162			
OY	219	GM-LPKLALIAA-----TGASSLCILITYSRNGMLGFVAMIFWALLGLYMPQRLPAPV		272			
		: :	: :				
Db	163	gyslarrvrrvaasflldgtgvtlmslaaap-----aafalmplrvlmlft-----	pw	210			
OY	273	--RRMLFPVVLGGLVAVLLAVALGCLPELRKRVLSIFVGRDSSNFRINWMLAVLQIID		330			
		: :	: :				
Db	211	gsrrlmltavaslafalplmctvpl-----	llaktqalf	245			
OY	331	RPML-----GIGGQNTAFNLVYPLXQARPTALSAVSPLYV-----	AVEGS	372			
		: :	: :				
Db	246	agwldyhwfgcfggvrrhvcfaafslfyyiknlwlf-alpa--lplawtvcrtllfscdwy		302			
OY	373	LLGLTFAFMLLLVTAVTAVRQVSRRLRDRNPQAF-----WLMASLAGLGMGLGHGLPDTY		427			
		: :	: :				
Db	303	llgv---vwmilaavlvlav-----ppqrfdnlvlllprla-----lfgaaqls-		344			
OY	428	LYRPEASTLMLWLCGATASF		447			
		: :	: :				
Db	345	lrrgaatlvwfgimaqglf		364			
RESULT 10							
AAV75522							
ID	AAV75522	standard; protein; 604 AA.					
XX	AAV75522;						
AC	AAV75522;						
XX	21-MAR-2000	(first entry)					
DT	XX						
XX	Neisseria meningitidis ORF 738	protein sequence SEQ ID NO:2518.					
DE	XX						
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; antigenic; diagnosis; immunogenic; infection; meningitis; septicemia						
KW	antibacterial; gene therapy.						
XX	Neisseria meningitidis.						
XX	OS						
XX	WO9557280-A2.						
XX	PD						
XX	11-NOV-1999.						
XX	FE						
XX	30-APR-1999;	99WO-US09346.					
PR	01-MAY-1998;	98US-0083758.					
PR	31-JUL-1998;	98US-0094869.					
PR	02-SEP-1998;	98US-0098994.					



Db 176 fligt-----gwtlmslaaaypaafalmjlp-----vlnff----- 207  
 QY 262 YMFQRLPAPW--RRMLFPVVLGGIVAVLLVAVLGLLEPLRVRVLSIFVGEDSSNNFRIN 319  
 Db 208 -----TPWGRIRMLTAAVASIAfAlpmtvypI----- 235  
 QY 320 VMLAVLQMTDRPWL-----GIGPCNTAFENLVYPLXQQAFTALSAVSYLEV--- 367  
 Db 236 -llakqpalfaqwdhvfqftggyvrhigtatlsfyiknllwfa-lpa--lpIavtvcv 291  
 QY 368 -----AVBGGLGLTAFAWMLLVTAVTAVRQVSRLRDRNPQAF-----WMLASLAGIA 416  
 Db 292 crrlftsdwgiIlgv---vmlavllvllav-----npqrfqdnllwllpIpa--- 335  
 QY 417 GMLGHGLFDVVLVLRPEASTLMWLCIGATIASF 447  
 Db 336 -lIfgaqIdS-lrrgaatvwmfIgaftglf 364

## RESULT 12

AAW90863

ID AAW90863 standard; Protein; 386 AA.

AC AAW90863;

DT 29-JUN-2000 (first entry)

DE Murine edg6 protein.

KW edg6; murine; G-coupled receptor; endothelial differentiation gene;  
 antiinflammatory; immunomodulatory; antimicrobial; antiallergic;  
 KX cytostatic; gene therapy; inflammation; autoimmune disease; allergy;  
 tumor; leukemia; lymphoma.

OS Mus sp.

PN DE19846979-A1.

PD 23-MAR-2000.

PP 13-OCT-1998; 98DE-1046979.

PR 11-SEP-1998; 98DE-1043240.

PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

PI Graeler M, Bernhardt G, Lipp M;

DR WPI: 2000-258069/23.

DR N-PSDB: AAX82790.

PT New human and murine G-coupled receptor EDG (endothelial  
 differentiation gene) 6, useful for modulating inflammatory and immune  
 reactions and for treatment of allergy or tumors

PS Claim 2; page 6; 12pp; German.

CC This invention describes novel human and murine G-coupled receptors EDG  
 CC (endothelial differentiation gene) 6 (I and II). The products of the  
 CC invention have antiinflammatory, immunomodulatory, antimicrobial,  
 CC antiallergic and cytostatic activity. (I) and (II) are involved in signal  
 CC transduction. (I), (II) and their fragments, variants and mutants or  
 CC binding partners, are used therapeutically to modulate the function of  
 CC blood and body cells, particularly for inhibition of acute and chronic  
 CC inflammation and to raise specific antibodies against them.  
 CC As a source of diagnostic oligonucleotides and for gene therapy.  
 CC Antibodies against (I) and (II) are useful for diagnosis and optionally  
 CC when coupled to therapeutic agents, toxins or other antibodies, to  
 CC modulate immune and inflammatory responses for example immunological  
 CC defects such as inflammation, infection, autoimmune diseases, allergy,  
 CC tumors, leukemia and lymphoma. This sequence represents the murine EDG6  
 CC protein described in the method of the invention.

SQ Sequence 386 AA:

Query Match 5.0%; Score 121; DB 21; Length 386;  
 Best Local Similarity 27.3%; Pred. No. 0.0006;  
 Matches 117; Conservative 33; Mismatches 98; Indels 180; Gaps 28;

QY 1 MYWQUTLTAHQPQOMGSSFLHRLFGSLRAMKRSOLLV-----WSEALG 47  
 Db 3 lctwstlv-----tpes-----chrlaas-----ghsllivlhyhsgrlaargsgedg 47  
 QY 48 GFLAVVGSAPFVPSSAIG-----LGLAIA-----AYMALISLTDIDLRQAT 91  
 Db 48 g--lgnlrg-----psvaagclvlenamvlaataiymrstrvyccllnltlsdl----- 96  
 QY 92 PIHMLVLVLYGWVDALATG-----LSPVRAALVGLAKTLVLLVPAARVLRNRLRSL 146  
 Db 97 -----ltglayvrvnlsgtrftglspvhwflreg-----llfmalaastf-----sl 139  
 QY 147 LFSVVYITSLFVSVYGLNQMITYGVEELATWD-RNSVADFTSRVSYIGNPMILAAYLP 205  
 Db 140 lft-----agerfatmrvnaesgatkstarygyciglcwllaail-- 178  
 QY 206 TTAFSAAALGVWRGMLPKLLAATGASSLC-----LILVYSGMVGFAVAMF---V 255  
 Db 179 -----gllpIrl-----gmcvcvafpcssllpIysk-gyvlIcavvfaIil 218  
 QY 256 WALLGLYWF-----QRLPA--PWRRL-----PPVVLGGLVAVLLVAVL 293  
 Db 219 valIslygaIfvrvrangqksprrparkrrllIntvImIvafvcvgplfglladIf 278  
 QY 294 G-----LEPLR---VRVLSIFVGRDSSNNFRIN-----VMLAVLQMT-QDRPW 333  
 Db 279 gsnvvaqeylrgmdwllalavf-----psaInplIysfrstevqavlaflccgclw 330  
 QY 334 LGI-GPGN 340  
 Db 331 lglrgpgd 338

## RESULT 13

AAV75523

ID AAV75523 standard; Protein; 604 AA.

AC AAV75523;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 738 protein sequence SRO ID NO:2520.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy.

OS Neisseria meningitidis.

PN W09957280-A2.

PD 11-NOV-1999.

PP 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

PA (CHIR ) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.



DB 254 riqnlyiangsvrqlsrmmhegqdlveklwlvavv|clavavamvr----- 302  
QY 230 TGASSLCLILTYSRGGLGFVAMIFWALLGLWFOPLPAPRR---WLFPPVILG----- 282  
DB 303 -----vragnpygagvamlinsllal|-----cspvswshmvwllpialglgas 345  
QY 283 -----GLVAVL-LVAVLGLPLEVRVLSIFVGRSDS-----SNFRRLNVWL 322  
DB 346 awngtrtapiatagvialliltitpmlfittfmpydseeyfpwlllqpsnagvyyvvl 405  
QY 323 AVLOMIDRRLWLGITGONAFN-----LVYPLYYQARF-----TALS 359  
DB 406 allivavivptv-lgsnkvavsgaeksspalvialafylfanivfkgnqpkall 464  
QY 360 AYSVPLEVAVEGGLGLTAFAMLLVTAVAVRQVSRRLDRNPQAFWLMASLAGLACL 419  
DB 465 qy---p|q-lmeq---r|ltdgellifeaasnqlvs-----lwllglnalaa--- 506  
QY 420 GRGLFPTVLYRPEASTLMMWL 439  
DB 507 -----latlwfl 514  
RESULT 15  
AAB43136  
ID AAB43136 standard; Protein; 535 AA.  
AC AAB43136;  
XX  
DT 08-FEB--2001 (first entry)  
XX  
DE Human ORFX ORF2900 polypeptide sequence SEQ ID NO:5800.  
XX  
KW Human: open reading frame; ORFX: detection; cycostatic; hepatotropic;  
KW vulnery; antiposrotic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;  
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
XX  
OS Homo sapiens.  
XX  
PN WO200058473-A2.  
XX  
PD 05-OCT--2000.  
XX  
PF 31-MAR--2000; 2000MO-US08621.  
XX  
PR 31-MAR--1999; 99US-0127607.  
PR 02-APR--1999; 99US-0127636.  
PR 05-APR--1999; 99US-0127728.  
PR 30-MAR--2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI: 2000-602362/57.  
DR N-PSDB; AAC77345.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
PS Claim 11; Page 4966-4968; 5507pp; English.

XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytosstatic; hepatotropic; vulnery;  
CC antiposrotic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
CC antihypoid; and antinaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
CC  
XX  
SQ Sequence 535 AA;

Query Match 4.9%; Score 118.5; DB 21; Length 535;  
Best Local Similarity 22.6%; Pred. No. 0.0017;  
Matches 89; Conservative 56; Mismatches 123; Indels 125; Gaps 19;

QY 1 MTWMTLTPFAHQ---PQMGHSSFLHRLFGS-----LRAMRSQDLWVSEAL-----GG 48  
DB 85 ltvvalcyaelgvtlpsggdyvkdifgglagflrlwlvavivlyptnqavaltfsn 144  
QY 49 FLAAVYGSAPFPSSALGLGLAAIAAAYWALLSLTDID---LRQATPIHML----- 96  
DB 145 yvlqplfpcc-fppesglrl-laal-----cllllwvccsvrwtvrgdiftgkllal 198  
QY 97 -VLHWGVDALATG---LSPVRA-----AALVGLA----- 122  
DB 199 allilmgivqlckgeyfwlepknafenfepdligvalafiqgsfaygwnflnyvteel 258  
QY 123 -----KLRLYLVLFAALARY-LRNRLLSLFSSVVTSLFVSVYGLNQW 166  
DB 255 vdpyrnlpraftisipylvtfyvanvaytamp--qellasnavaVcfgeklligwaw 316  
QY 167 IYGV-EIATWYDRNSVADFTSRVYSYLGNPLLAAYLVPTTAFSAAIIGWRGMLPKLL 225  
DB 317 implivalstfgvngslfssrlf-----fagareghlpavl 354  
QY 226 -----AIAATGASSLCLILT---YSRGGLGFVAMIF---VMALLGLWFOPL 266  
DB 355 amihvkrctipalllftclstlmvtsdmyllinygfyfnylygvvavagivlrwkkp 414  
QY 267 RLAPAPRR-RMLEPPVVLGIVAVLVAVLGLLEPL 298  
DB 415 dlpripkinllfpilyllfwallvfwslwsepv 447

Search completed: January 10, 2002, 23:04:03  
Job time: 2290 sec

1

1

1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2002, 23:04:53 ; Search time 33.48 seconds  
(without alignments)  
1062.530 Million cell updates/sec

Title: US-09-887-038-3  
Perfect score: 2426  
Sequence: 1 MTWQTLTFAHYOPQWGH...WQPPSKQLPPEAHSDEKM 467

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pIR\_68: \*  
2: pIR1: \*  
3: pIR2: \*  
4: pIR3: \*  
5: pIR4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1132	46.7	474	2	S75464
2	206	8.5	438	2	A82262
3	201.5	8.3	387	2	S74522
4	195	8.0	443	2	D82847
5	180	7.4	478	2	A83368
6	169.5	7.0	384	2	H75597
7	166.5	6.9	597	2	H82328
8	159	6.6	401	2	S77086
9	151.5	6.2	461	2	D83022
10	151.5	6.2	461	2	G83602
11	139.5	5.8	388	2	G75375
12	138.5	5.7	402	2	E86038
13	138	5.7	515	2	G75267
14	137.5	5.7	443	2	F83759
15	137.5	5.7	474	2	H75389
16	134.5	5.5	660	2	A64739
17	134	5.5	335	2	G82224
18	134	5.5	402	2	A83398
19	133.5	5.5	585	2	S74477
20	132.5	5.5	660	2	E85499
21	131	5.4	524	2	T02499
22	130	5.4	649	2	F83779
23	129	5.3	487	2	B72755
24	128.5	5.3	530	2	G83631
25	128.5	5.3	566	2	H84203
26	127.5	5.3	437	2	D83329
27	127	5.2	591	2	F70871
28	127	5.2	709	2	F75584
29	126	5.2	524	2	T03507

30	126	5.2	776	2	A84178	cationic amino aci
31	125.5	5.2	404	2	G83322	hypothetical prote
32	125.5	5.2	787	2	H70374	NADH dehydrogenase
33	125	5.2	403	2	C83422	nitrate transpor
34	125	5.2	486	2	F83315	NADH dehydrogenase
35	124.5	5.1	328	2	D82411	conserved hypotet
36	124.5	5.1	553	2	C81024	hypothetical prote
37	124.5	5.1	643	2	G75599	probable drug tran
38	124	5.1	363	2	A72771	hypothetical prote
39	123.5	5.1	382	2	D75541	probable Cynx-rela
40	123.5	5.1	500	2	F83418	sodium/proton anti
41	123.5	5.1	501	2	G82822	NADH-ubiquinone ox
42	123.5	5.1	539	2	H82994	probable permease
43	123.5	5.1	604	2	A81181	hypothetical prote
44	122.5	5.0	404	2	B41317	O-antigen ligase c
45	122.5	5.0	558	2	F81967	probable inner mem

## ALIGNMENTS

RESULT	1	46.7%	Score 1132	DB 2	Length 474
S75464	hypothetical protein slr1515 - Synecocystis sp. (strain PCC 6803)	Best Local Similarity	47.5%	Pred. No. 2.8e-73	
C:Species: Synecocystis sp.		Matches 223	Conservative 79	Mismatches 163	Indels 4
A:Variety: PCC 6803					Gaps 2
C:Date: 25-Apr-1997	#sequence_revision 25-Apr-1997	#text_change 08-Oct-1999			
C:Accession: S75464					
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu					
DNA Res. 3, 109-136, 1996					
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys					
S:Reference number: S74322; MUID:97061201					
A:Accession: S75464					
A:Status: preliminary					
A:Molecule type: DNA					
A:Residues: 1-474 <K&N>					
A:Cross-references: EMBL:ID90911; GB:AB001339; NID:q1653083; PIDN:BA018025.1; PTD:d101					
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996					
Query Match					
1	MTWQTLTFAHYOPQWGHSLHRLFCGLRAMRASSOLVWSEALGFLAVVYGSAPF 60				
2	ISIMRSLMGFGFSPQEMGRGSVLRHLYVCGQSMVQASVLMHFEALGVALVAILFIAPF 64				
3	VPSSALGIGLAAIAYVALSLTDIDLQATPTIMVLLVGVDAALATGSLSPVRAALVG 120				
4	TSTWLGIFMLLGFALVLPFADQPGKLRPHIVLVFAVYCISALVGVSPVMAAASG 124				
5	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
6	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
7	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
8	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
9	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
10	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
11	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
12	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
13	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
14	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
15	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
16	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
17	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
18	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
19	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
20	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
21	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
22	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
23	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
24	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
25	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
26	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
27	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
28	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				
29	LAKLIVLVFALAAVRLNRLSLFSVVVITSLFVSVMQWVGYVEALATWVDRN 180				

OY 419 LGHGLFDVLYRPREASTIMWICGAIASFQPOPSKOLPPRAESDEKM 467  
 Db 423 LVHGVMDVWYRPPVSTLWMLLVIAVSQMASQAARLEASKRENEKPL 471

## RESULT 2

A82262  
 conserved hypothetical protein VC0931 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: A82262  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, R.  
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: A82262

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-438 <HEI>

A:Cross-references: GB:AE004176; GB:AE03852; NID:9655385; PIDN:AAF94093.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0931

A:Map position: 1

Query Match 8.5%; Score 206; DB 2; Length 438;  
 Best Local Similarity 24.7%; Pred. No. 1.7e-07;

Matches 120; Conservative 68; Mismatches 200; Indels 98; Gaps 19;

OY 5 OTLTFAHQPOQWGHSSFLHRL-----GSLRAMRASSQLWVS--PALGGFLAAVYG 56  
 Db 6 KTLFPA-----FLFVFWLPDPLGSRNPW-----ANSLNITYAFLL----- 42  
 OY 57 SAFPYSSAAG-IGIAIAAAYWALSLTDIDLRQATPIHMLVLWYGVALATGLSPYRA 115  
 Db 43 -----LGGFLPRAACMLAIGRAKVL--IPALFTLMSWQSWSLGSSQOT 90  
 OY 116 AALVGLAKITLYLVFALAARVLRNRLSLFSYVITSLFVSYG----- 162  
 Db 91 LVFISAVSLSHLYQICVLAALVDPHRLKMLATMTMIASGICQAFYAGVILLLEQTSPE 150  
 OY 163 -----LNGMIVGEELATWDRNSVADFTSRYSYLGPNLLAAYLVPTAASAAIGYWR 218  
 Db 151 FSLPLNQAASG-----SFYVNHHLANFL-MLNLCIGFGLITAEILNHQTO-----GWR 197  
 OY 219 GMLPKLLAIAATGAS-----SLCLILYSSRGMLGEYAMIFVWALLGLYWFQPLP 269  
 Db 198 NTLKLLAILLSDKAFIRLGLIVINIALVLTFRSRMGNTAFVVALSIGILLILEFKSR-- 255  
 OY 270 APMRWLEFPVYLGIVAVILVAVLGLLEPLRVRVLSIFGRESSNNFRINWMLAVLQMIQ 329  
 Db 256 -PKSLYLLIVLSLEIVDALVSNMFGLDKVRQR-----LAETSLQWESKDDVVRDLNLIQ 309  
 OY 330 DRPWLIGIPGNTAFMLVPLYOQARFTAL--SAYSVPLEAVVEGGLGLLTFAMLLVTA 387  
 Db 310 EHPPLGFG--GNGTFYSTPSPYQSGEVALFYDHAHNDYQFTLESLLINTSVFGIIVLCA 367  
 OY 388 VYAVKQVSRLRDRNPQAFWMLASLAGLAGMIGLFTVLYRPREASTIMWICGAIASF 447  
 Db 368 WRAL-SAFENRFTNKTMRGIGL-GSLMAITIGMLHMSVDFIQAP-ATALYFLFCLLIANW 424  
 OY 448 WQOPS 453  
 Db 425 SMTLPT 430

RESULT 3  
 S74522  
 hypothetical protein slr0728 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S74522

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.; K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74522

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-387 <KAN>

A:Cross-references: EMBL:090899; GB:AB001339; NID:q1651650; PIDN:BA016674.1; PID:dl01

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

Query Match 8.3%; Score 201.5; DB 2; Length 387;  
 Best Local Similarity 25.9%; Pred. No. 3.1e-07;

Matches 103; Conservative 54; Mismatches 179; Indels 61; Gaps 17;

OY 95 WLVLWYGVDAATGLSPVRAALVGLAKLTLVLVFAALAVLRNRLSLFSYVIT 154  
 Db 2 WAALTLMFL-LFTTTHDRGKALLGLANFVYFIYAFASQYICQKOLNTLMLITAN 59  
 OY 155 SLFVSYGLNOMIYGEELATW---VDRNSVADP--TSRYSYLGPNLLAALVPTTF 209  
 Db 60 TVVILVITIGFQ-VYGGWATPNNMLVAGTGNLVAGRGPRGRSSLLMYANLFSALLWFPPL 118  
 OY 210 S-AAAGVNRGM-----LPKLL--AIAATGASSLCLILYSSRGMLGEYAMIFW 256  
 Db 119 SLGLIISVRRMHTAPNSFNPLPLVWALLVACILEALVLTGSSAM-GIGLLIGTA 177  
 OY 257 ALLGLWFPRLPAPWRRLFPVVLGGLVAVLLVA---VLGDEPLRVYLSITVGR--- 309  
 Db 178 YAIYLSWY-----WLV-ALAGGATAMVLMASGPGKPEPLQIYKXYWGRSLDE 226  
 OY 310 ---EDSSNNRINWLAIVLQIDRPWLIGPGNTAFNLYPPLYOQARFTALS-AVSPL 365  
 Db 227 LYPDRYTRALSTQWQPSWDFLDQPLFGGGLNFT-----PLYOANMYWIGHPHNVL 261  
 OY 366 EVAVEGGLG---LTFAMLLVTVATAVAVRQYSRLRD---RNQAFWMLASLAGLAG 417  
 Db 282 MMIGETGLIGTALMLGAVGFLAOGVLLI-----HLSRGGFRRRSQHLLLSYIAFAA 337  
 OY 418 MLGHLFDVLYRPREASTIMWICGAIASFQPOPSK 454  
 Db 338 LCIVNLPDVTIFDMRNVLGMIFLAATAGVSORVGTG 374

## RESULT 4

D82847  
 membrane protein xF0103 [Imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: D82847

R:anonymous, The Xylella fastidiosa Consortium of the organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82847

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-443 <SIM>

A:Cross-references: GB:AE003864; GB:AE003849; NID:9104879; PIDN:AAF82916.1; GSPDB:GN

A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre

as-Heto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.





[illegible]

```

RESULT 7
H82328
hypothetical protein VC0393 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: H82328
R:Heldeberg, J.F.; Eisen, J.A.; Nelson, W.C.; Claydon, R.A.; Gwin, M.L.; Dodson, R.J.;
charson, D.; Ermolaeva, M.D.; Vamathavan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, E.
J. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nucleo 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20405833
A:Accession: H82328
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-597 <HEF>
A:Cross-references: GB:AE004127; GB:AE003852; NID:g9654808; PUDN:AA93566.1; GSPDB:GN000
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0393
A:Map position: 1

```

Query Match	6.9%	Score 166.5	DB 2	Length 597
Best Local Similarity	24.3%	Pred No. 0.00015		
Matches 109	Conservative	60	Mismatches 185	Indels 95
			Gaps	22

  

QY	49	FLAAVYGSAP--FVPS-SALGIGIAAIAAYALLSLT-----DIDRQATPIHIVLL	99
		: : : :   :   :	
DB	29	FALAAVYLLAMFEFMPNPGAGLALSFNTVWATISITAIAGLYOLANNOALRYEKLTIG	88
		: : : :   :   :	
QY	100	YNGVDALATGSLPV-----RAAIVGL-AKLTLYLVFALAAVLRNRLSLFS	149
		: : : :   :   :	
DB	89	LMVSCLLLT--LPITYSNADWFHFLGRLGIMAGFTLLFLVLOO--RESNKHQRLLMW-	143
		: : : :   :   :	
QY	150	VAVITSLFVYGLNQMIQVEELATWDRNSVADF--TSRVSXYGLNPNLAIVLPT	206
		: : : :   :   :	
DB	144	ILLGSEIVAIIGLIQYF-----WLEGNPGVDIEANRPYGLFOQPNVMASPLATNG	194
		: : : :   :   :	
QY	207	TAFSAATGVWRGWL-----PKLIAATG-ASSICLILITYSRGNLCGFVAMI	253
		: : : :   :   :	
DB	195	LVLS-----GYLLARPEKYDSHWSKIGILYATPILTAIPLAVLASRTGMASLSTSI	246
		: : : :   :   :	
QY	254	FWALLIGLVQFQPRDPAAPWRMFLPVVLCGLAAVLLVAVLGEPYLRVRLSYFVGREDSS	313

[illegible]

RESULT 8  
S77086  
hypothetical protein sl10737 - *Synechocystis* sp. (strain PCC 6803)  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S77086  
O:Kaneoko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, J.; Kato, R.; Okunura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *synechocystis* S.  
S  
A:Reference number: S74322; MUID:97061201  
A:Accession: S77086  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-861 <KAN>  
A:Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BA10778.1; PID:g100  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Superfamily: *Synechocystis* hypothetical protein sl10737

```

Query Match Similarity      6.6%: Score 159: DB 2: Length 861:
Best Local Similarity      22.6%: Pred. No. 0.00073:
Matches 112: Conservative 63: Mismatches 145: Indels 176: Gaps 30:

QY      48  GFLTAIVYGSAPEWESS-ALGLG-----LAAIAAYW-----ALLSTIDLR 88
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      27  GLTFLATLGLFTLLPENGSLMGWPMWPFVWQTMETLALMCIMQIGONKRLISL----- 80
QY      89  QATPIHHMLVLLYWGVDALATGLSPYRA-----AAVLGLAKTLLYLVLPALA 134
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      81  -GLGEGDWMLVAL-----ATAAMLVYGAEFPQOARVYGCFAFLAGLGLRSMW----- 129
QY      135  ARVLNRP--RLRSL-----FSVVVITSLFVSAYGL-----NOMIYG 169
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      130  -KVQNPPLVYSINIKLTFEGYGVFAFIVISLLMFQTQLIPWQSAIAREMGINKRFTFG 188
QY      170  VEELATWMDRNSVADFTSRVSYLGNPLLAAYLPTPAFSAALIGW-----RGWPKL 224
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      169  ILELDNMP-----IGHQNVAYGVLV--LILPLSLVILINLNGKRKRWMSV 232
QY      225  LAIAATGASSICLI---LTYSRGGWLGFWAMIFWAILGLYEFQPRLPAPRRMLFPVVL 281
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      233  -----ALSLGLIDFYTTSRSGLLG--LATLLILVILIGILRL---QLPRRWL----- 276
QY      282  GGLVAVLLVAVLGL-EPLRVRVLSFEVGRESSN-----NFRINWMLAVLQMTODR 331
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      277  -GL-GLTAAIVAGIETINDRLITSETGIMGGAGQAFAYLLINEIG-W-----RMGSAH 329
QY      332  PWLGISGPENTAFNLYPYL-----QQAREFALSAVSYLEVAVEGGLGLTFAFWMLLV 385
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      330  PWTGTGLGNVP--LQYOLXRPVWAGRESEF-IYOLHSTPADLFAELCY-----WGILII 379
QY      386  TAVTAVROVSRRLRRDNPQ-----AFWLMASTLAGLGMIGHLEFDTVL 428
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      380  FSLITIGIIMQIRSLIKQNTVTRKNDVDSNVYKILMTLTLT-----ALLAYVSTLTD 434

```

QY 429 YRPEASTLWMLCIGAI 444  
 Db 435 YQLDN-----VCISGV 445

RESULT 9  
 D83022  
 hypothetical protein PA4999 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: D83022  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,  
 .; Lory, S.; Olson, M.V.  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: D83022  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-401 <STO>  
 A:Cross-references: GB:AE004912; GB:AE004091; NID:g9951274; PIDN:AA08384.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4999

Query Match 6.2%; Score 151.5; DB 2; Length 401;  
 Best Local Similarity 24.3%; Pred. No. 0.0011;  
 Matches 113; Conservative 53; Mismatches 156; Indels 143; Gaps 28;

QY 17 W-GHSSTLHRLFGSLRANRASSQLVWSSEALGFLAVVYGSAPVPSSAL 75  
 DB 36 WVGGRSDYHRLFTLLA--APTLLV--LDPRLRLPTGSPLEI-----AFLA 83  
 QY 76 YMALSLTDIDLROATPIHMLVLLVYGVDAATGLSPVRAALVGLAKLTFL--LVFAL 133  
 DB 84 Y-WMLSLT-----WSTPENSTG-----SLKRPILYLLFPCA 115  
 QY 134 AARYLRNP-RLR-----SLFSVYVITSLFVSYGLNOMITVEELATWDRNSVADFTS 187  
 DB 116 AILALAPRLAKRTWMLAALGAVISAATILRY-----W-DANPL----- 155  
 QY 188 RV--YSYLRNPILAAVLYPTAFSAATAGVRCMLP-----KLIAIATGASSLCITLT 240  
 DB 156 RLTYGTGLYNP-LLSAHY--GAFTALMLAYMOSRPIALPPLISLALGG--LLIAT 209  
 QY 241 YSRGWLGFYAMLFVALLGLVWFOPRLPAWRRLFPVVLGVLAVLLVAVLGLLEPLRV 300  
 DB 210 GSRPLVGLTAAAL-WMLVLA-----GDRKKALLALAL-ALAGALLGYILYEVITYQ 258  
 QY 301 RYLSIFVGRDSSNNFRINWLAVLQMIODRMLIGIPGNTAFNLVYLYQOARTALS- 359  
 DB 259 RGAAS-----FRPEIWDALRQISEHFWLGHG-----YDHPKRIYLSN 295  
 QY 360 -----AYSPLEVAEGGLGLTAFAMLLVYTAAYAVOVSRLRRDRNPQAF-----WLM 409  
 DB 296 GMLADHNIELGVLFGGITG-----LLWVAIYALAFGSRMKRRSPAVLLASTLV 349  
 QY 410 ASLAGLAGMLGHGLFTVLYLRPEASTLV---WLCIGALASFQWQ 451  
 DB 350 FGLAA-----GLTGSNAFLPRPKHMLVIMPALVIALMTIQ 387

RESULT 10  
 G83602  
 hypothetical protein PA0345 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: G83602  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,

.; Lory, S.; Olson, M.V.  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337

A:Accession: G83602  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-461 <STO>  
 A:Cross-references: GB:AE004472; GB:AE004091; NID:g9946188; PIDN:AA03734.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA0345  
 C:Superfamily: Neisseria meningitidis hypothetical protein NM1645

Query Match 6.2%; Score 151.5; DB 2; Length 461;  
 Best Local Similarity 23.7%; Pred. No. 0.0013;  
 Matches 120; Conservative 60; Mismatches 155; Indels 171; Gaps 29;

QY 7 LTFPAHYPOQWCHSSEFLHRLFGSLRANRASSQLVWSSEALGFLAVVYGSAPVPSSAL 66  
 DB 49 LTRAHMLGRREG-----LIDALRTWQCSRLA-----ALLVIALASGAGIAPAL 95  
 QY 67 GLGLAAIAYVALLSTPDIDLROATPIHMLVLLVYGVDAATGLSPVRAALVGLAKLTFL 126  
 DB 96 GDGQRPVWVFVWLASLIG-----LHLTLTLMALGLLAG--EAGAGLG--RLML 141  
 QY 127 YLLVFAARVLRNPRLSLFSVYVIT-----SLFVSYGLNOMITVEELATWWD 178  
 DB 142 W-----LSGLIARORAAHMLPALVLLGRRRLARWGLGALVHGL--WLGILTLTALML- 193  
 QY 179 RNSVADFTSRVSYTYGNINLAA--YLPPTAFSAALIGVRCMLPKILA-----IA 228  
 DB 194 ---LGLATRRYGFWEETITLSDPTALTQALGA-----LPALLGFPPLDAELIR 241  
 QY 229 ATGASSLCILITYSRGMIGF-VAMIFW-----ALLGL--YWFOPRLPAWRRLFPVY 280  
 DB 242 ASGDAAAL--SEAAHMAAGVGLVGLVPLPALLGLCLM-----RMKRGIAHLD 292  
 QY 281 LSGLYAVLLVAVLGLFPLRVVL--SIFVGRDSSNNFRINWML-----AV 324  
 DB 293 LD-----LDDPGYSLLRERLMPASERLGVSDAAP-----WLPFGQGGQGEAAV 340  
 QY 325 IQMIO---DRPMLGIPGNTAFNLVYLYQOARTALSAYSPLEVAVEGGLT--GLTAF 379  
 DB 341 LVAVELDDRPP--PRKLA-----EGVADAGILDDGGQQR 373  
 QY 380 AMLLVYTAAYAVOVSRLRRDRNPQAFWMAASLAGLGMGLHGLFTVLYLRPEASTLML 439  
 DB 374 RLLEQUTRYPPARUALACDPRRSPD-----RGTIALIGE-----IARCAASTYRWL 419  
 QY 440 CIGAIASFQWQPSKOLPP-EAEHSD 464  
 DB 420 L-----QAPPGALDSD 431

RESULT 11  
 G75375  
 hypothetical protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: G75375  
 R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J  
 .; M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: G75375  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-388 <WHI>  
 A:Cross-references: GB:AE002004; GB:AE000513; NID:g6459366; PIDN:AA11168.1; PID:g645

```
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1602
A;Map position: 1
```

Query Match	5.88;	Score 139.5;	DB 2;	Length 388;
Best Local Similarity	28.48;	Pred. No. 0.0079;		
Matches	84;	Conservative	24;	Mismatches 81;
				Indels 107;
				Gaps 20;

```

OY      20  SSELHLFSLSTLAWMAASQILWSEALGCLTAVVYGSAPEFPSSALLGILALAAVWAL  79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      146  ASFFAALISGLRG-----RWQALAAALALAAVLFHTFHPDHC-GIGVS-----  189
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      80  LSLTIDLRLQATPIHML---VLLWGYDALATGSPVRAALVGLAKLTLYLFWLFAALAR  136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      190  -----ESANPVOYLNDALLSAMGRGL--LSVPTTALV-----LLGALAAAR  229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      137  VL--RNPRLRLFLFSV-VYIITSFVS-----VYGLNOMITGVSELA  174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      230  PLOQKNRPAPLLLLGLGAVLTALGMAASGRULPFSKALMTPTPYLLYSAGLSTGLI--LA  287
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      175  TWVDNRNSVADF-----TSRKYVSYLGNP--MLLAAYLVPTTAESAAAIQWV-----RGWL  221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      288  CWV---VADSGWLPQGRKLILAPLITGRNALIAGVLP-----ILIKWITLIDMQVGT  337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      222  PKLLAIATGASSLCILITYSR-----GCW---LGFYAMFVWALLGLTYFQPR  267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      338  GKSSQISAS-----LLEMASSSFGPYGCGWYITLGLV--FAVW--LGLAMMAR  382

```

RESULT 12  
E86038  
probable LPS biosynthesis enzyme waal [imported] - Escherichia coli (strain 0157:H7)

C:Species: *Salmonella* *coar*  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: E86038  
R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
H.L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Mature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Accession: E86038  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-402 <STO>  
A:Cross-references: GS:AE005174; NID:g12518378; PION:AA658769.1; GSPDB:GN00145; UWGP:255  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: wael

Query Match	5.7%;	Score 138.5;	DB 2;	Length 402;
Best Local Similarity	21.8%;	Pred. No. 0.0096;		
Matches	85;	Conservative	67;	Mismatches 151;
			Indels	87;
			Gaps	19;

```

QY      26  LFGSL--RAWRASSQLLVWSEALGCELLAVVYSGAPVPSBSALGLGIAIAIAY---WAL 73
      5  LFFSLKKRN-----TAVNNRAL-VFLEITTY-----FLGIDTRYKHLIVIL 45
Db      80  LSLDIDILROATPIHMLVY---LWYGDALATGL-----SPVRNAALVYGLAKTL-LYL 128
      46  MTTTITTYLCKRKHKHLSFKETFLGSAVILITALLSLLSQSDNAGASKETFAIIENT 105
QY      129  LVFALAARVLNRDLRSLLESVVVITSLEVSVGL-----NQMIVGEELATWVD 178
      106  LLCTIAPILLRDEKREDEVERKIVFFS--FISALGLRCFSELITIVYDYOGGIMPADYRH 163
Db      179  RNSVADITSVSVYSGNPVLAAVLYPTTPASAAAGVWGMGLPKLAIATAAGASCLDI 238
      164  R-STSD--SWVLEF---PALLMMLITKAKR-----ISVYLVSTFIYILI 203
QY      239  L-TYRSGWGLGFVAMIFVALLGLVWFQPRILPAWRIRMLFPVYVGLGVAVILLVAVGLEP 297

```

```

Db      204  LGTSLRGAWLSVILGILWILM-----FKQWKL-LLVGVMAIIALSVIFTHK 250
          | | | | | : : : | :
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 251 EMTAKLTLYKLOQTSSRYRANGTQGSALDILEENVGIGYGNVAKYKYYNKRVIDPEW 310

```

QY      354 RF-TALSAYSVPLEVAVEGGIGTFAFWL 362
        | :: :: | : |||| : |
DB      311 TFRQSIGNHNALFTINFTGLGGLVSLMML 340

```

RESULT 13  
G75267  
ABC transporter, permease protein, CysW family - *Deinococcus radiodurans* (strain R1)

C:Species: Dermococcus leucodermis  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: G75267  
R:White, O.; Eelsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A,Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* RI.  
A,Reference number: A75250; MUID:20036896

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-515 <WHI>  
A:Cross-references: GH:AE002078; GH:AE000513; NID:66460306; PIDN:AP12029.1; PID:9646  
A:Experimental source: Strain R1  
C:Genetics:  
A:Gene: DR248  
A:Map position: 1  
C:Superfamily: sfus protein

Query Match	5.7%;	Score 138;	DB 2;	length 515;
Best Local Similarity	23.6%;	Pred. No. 0.013;		
Matches 102;	Conservative 51;	Mismatches 154;	Indels 126;	Gaps 20;

[illegible]

```

Oy      209  FSAALGCVWRGRLPKLLIAATGAS---SLCLILYSRGNLGFVAMIFVALLG----- 260
          :  ||  ||  :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      251  LTRGAGVATGCLPRARGALMGITLLGLTLLVCFAP-----LLAVVRGLVGSAGPT 304

```

Db 305 LTYW-----RGVAVNDEQTPLLGGNTLRFGALALLGNTAL-----CG 340

Db 341 LVALGAWLARSRYLDLISLP-LMSPVSLAVGYLATPYVAATLPMIAATLLALPLL 399

Db 400 VRSLLPALRAIPRLFEARPL--GASPLAHRRTVTLPLTLPALRGCAALATVIGEEFC 457

```
QY      424 FDTVLYPEASTL 436
          1111:11
Db      458 ATLVLRPEWATL 470
```

RESULT 14  
F83759  
cytochrome d (bd-type) ubiquinol oxidase subunit I BH0878 [imported] - *Bacillus halodurans*  
C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: F83759  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: A83650; M01D:20263314  
A:Accession: F83759  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-443 <STO>  
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04597.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH0878  
C:Superfamily: cytochrome d complex terminal oxidase chain I

Query Match 5.7%; Score 137.5; DB 2; Length 443;  
Best Local Similarity 21.5%; Pred. No. 0.012;  
Matches 91; Conservative 61; Mismatches 129; Indels 143; Gaps 20;

QY 44 EALGGLAVVYGSAPFVPSAL-GIGLAAIAAYMALSTLTDIDLRQATPIHMLVLLYNG 102  
DB 105 EAL-FLSIYVADRLTALRLVSVFVALGATASAVLITTDHAMMNTPRGFTT----- 157  
QY 103 VDALATGLSP-----VRAALVGLAKLTLYLLVLAARVLRNRL----- 143  
DB 158 VDAAVTVNRPLEAVFNPSTIYVTAMHVGTAYMGAIVASVAAKLFRCNLSREASVHR 217  
QY 144 RSLFESVYV--ITSLFVSYYG-----LNQW-----ITYGEELATWDRNSVADFTSRVVS 191  
DB 218 KGMFPLAIGLMSITTAINGHDTAKMLHQVLPVKIAAEGLETDQNAPLAIF----- 271  
QY 192 YLGNPMLAAYLVPTTAFSAAGVWRGMLPKLAIATGASSICL--ITFSRGGM--- 246  
DB 272 --GTPSEAGRYIGLE-----IPGLSLMALGCTTDGVYKGLYDPRNRMPPL 317  
QY 247 -----IGFVAMITVWALLGLYRQPLPAFRWRMLFPPVVLGLVAVLAVLGL 295  
DB 318 FVHTLFNMVYIGFSLIGLAVACWGFPPFKRPPF--KWL-----LAGLIASGLAMIGI 371  
QY 296 EPLRVAVLSIFRGREDSNNFRINWVLAVLQMIQDRPMIGIGBNATFNVLPVQOARF 355  
DB 372 E-----TGM--TFSGTGRQPM----- 385  
QY 356 TALSAVSPL--EVAVEGGLG-----LTAFAMLLVTAATAVROVSRLRDRNPQAFW 407  
DB 386 ---TIYGIQLTSEATQSGNLGFLFTLITLAVALLVITAL-----VMAFFYHRNPVSNE 437  
QY 408 LMAS 411  
DB 438 WMS 441

RESULT 15  
H75389  
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 [similarity] - *Deinococcus radiodurans*  
C:Species: *Deinococcus radiodurans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 05-May-2000  
C:Accession: H75389  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: H75389  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-474 <WHI>  
A:Cross-references: GB:AE001993; GB:AE000513; NID:96459244; PIDN:AAFL1056.1; PID:9645  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRI493  
A:Map position: 1  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4  
C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 5.7%; Score 137.5; DB 2; Length 474;  
Best Local Similarity 23.7%; Pred. No. 0.013;  
Matches 114; Conservative 54; Mismatches 162; Indels 151; Gaps 26;

QY 42 WSEALGGLAVVYGS-----PEFVS-----SALGIGLAAIAAYWA 78  
DB 23 WREEVAFPPAAATLGAFLMLRGLGEGSLVLPWPPLGTVAVELNGVLAFAVTAEMT 82  
QY 79 LSLTDIDLRQATPIHMLVLLYGVDAALATGISPYRAAALVGLAKLTLYLVFALAARVL 138  
DB 83 LAAVYTAARVRNPGMPLAM---VLAMETGLGIFAA---RDLILEYFEFMA----- 129  
QY 139 RNPRLSLFVSVVITSLFVSYYGLNOMLYGEELATWDRNSVADFTSRV--YSTLGNPN 197  
DB 130 -----LISLMLAVYGGPHRRALVKFAATYLLGSLPMLLSPIGKMLGGS- 176  
QY 198 LLAATLVPTTASAAATGWRGMLPKLAIATGASSICLITITSRGGMGLFVAMIFWA 257  
DB 177 -----PTFA-----LPDLLAHVPYGVAAQTML-----YLGFLAAMAV-- 207  
QY 258 LGLYWFQRLPA--PMRML-----FPVYLGIVAVLVAVLGLBPLRVRLSI 305  
DB 208 -----KLPMLPLHMLPDPFHEONHESGIPDVWGLTKV-----GGYGIFFYGLTL 252  
QY 306 FVGRDSSNNFR--INWVLAVIOMIQDRPMLGIGPNTFENVLPYQOARFALSAYSVP 364  
DB 253 F---PDASLELRPVLMAALFAATLV--AAWIAFGQNNMRLILAVGLSHMGVAGLGFSMN 308  
QY 365 LEVAVEGGLGTAAMLLVTAATVAVROV-----SRLRRDRNPQ-----AFWL 408  
DB 309 -ETAVIGAMV--LLAQONLYTGALFMAVGMLOERVSGSLDTRVGGVNNQAGALGILTLAM- 365  
QY 409 MASLA--GLAGMLGGLPDTVLYRPEASTLMWL-----CIGATA-----SFMPO 451  
DB 366 FASIAVPGIAGFVG--FSILGATQVSP--WLTPIAGLTITIAAAVATLTAPOKTFMEGR 421  
QY 452 P 452  
DB 422 P 422

Search completed: January 10, 2002, 23:11:14  
Job time: 381 sec

---







```

QY 61 VPSSALG-----LGLAIAAYMALSLTDIDLRQATPIHMLVLLYGVDAALATGLSP 112
D 210 RPLTIMGIDGVARNGLGLSLARLALSLAVY-----ISALL----- 247
OY 113 VRAAALVGLAKLKLVLVLFALARVLRNRLSLFVSVVITSLFVSYVGLNQMIVGEE 172
D 248 VNAVGIITGFLPAPLAKMLKARRLL-PRL-MLASLIGALILWLS-----DQ 293
OY 173 LATVDFNSVADFTSRVSYVSLGNPNLLAAYLVPTTAFSAALGVWRGW-LPKLLAIAATG 231
D 294 IIMLFRVMEVSTGVTALIGAPLL-----MLPRLKSTISAPD 333
OY 232 -----ASSCLILTVS-RGCVLGFVAMIFWAL-----LGLYWPQRLPAPWRMLPPV 279
D 334 MKVNDRAERQHYALFALAGVLLMAVVALSPGRDAGHTWASGALLEDMWRMRPR 393
OY 280 VLGLVAVLIVAVLGLPELRVRLSLFVGRSDSNFRINWMLAVLQIDRPW-----L 334
D 394 IMAALFAGVMLAVAG-----CITQRLTGPMASPEVL 425
OY 335 GIGPGTAFNLVPLVQOARFALASVPLEVAEGGLLGTAFAML-----LTVATV- 389
D 426 GISSG-AAGGVLMLE-----LVPGNAGWMLLPAGSLGAAYTL 462
OY 390 AVQVSRLLRRDNPQAFWIMASLAGLMIGHLPDVLVYRPEAS-----TLMWL 439
D 463 LIIIMAGRGGSFPHRL-----LAGMALSTAFMLLMQASGDRMAQVLTW 512

```

## RESULT 3

```

ID REFL_SALTY STANDARD: PRT: 404 AA.
AC P26471;
DT 01-AUG-1992 (rel. 23, Created)
DT 01-AUG-1992 (rel. 23, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE O-ANTIGEN LIGASE.
GN REFL OR MAAL OR REBT.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=92041612; PubMed=1657881;
RA MacLachlan P.R., Kadam S.K., Sanderson K.E.;
RT "Cloning, characterization, and DNA sequence of the rfaL region for
RT lipopolysaccharide synthesis in Salmonella typhimurium LT2.";
RL J. Bacteriol. 173:7151-7163(1991).
CC -1- FUNCTION: ADDS THE O-ANTIGEN ON THE GLUCOSE(II) GROUP OF LPS.
CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: SHOWS ALMOST NO SIMILARITY TO E.COLI REFL.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M73826; AAA27206.1; -
DR PIR: B41317; B41317.
DR StyGene: SG10340; rfaL.
KW Lipopolysaccharide biosynthesis; Ligase; Transmembrane.
FT TRANSMEM 66 84 POTENTIAL.
FT TRANSMEM 184 203 POTENTIAL.
FT TRANSMEM 227 244 POTENTIAL.
SO SEQUENCE 404 AA; 46034 MW; 1B84F7B3C757D2AB CRC64;

```

Query Match 5.0%; Score 122.5; DB 1; Length 404;  
 Best Local Similarity 20.8%; Pred. No. 0.39;  
 Matches 68; Conservative 63; Mismatches 123; Indels 73; Gaps 16;

```

OY 129 LVFALARVLRNRLSLFVSVVITSLFVSYVGLN-----QWLYGEELATWDRNSVADF 185
D 105 LVTTLFVLVLLKQETKEVAKIYFES-FLTSLGRLCAESLITIEDY-----NKGIWPF 157
OY 186 TSVRYSTLGNPNLLAAYLVPTTAFSAALGVW-RGWLPEKLLAIATGASSCLILTVYSR 243
D 158 ISVAHRHMSDS--WVLEFP-----ALINIMLFKRNALKIVFLVLSALYLFILIGLSR 208
OY 244 GGMVGFVAMIFWALLGLVWFOPLPAPWRMLFPVVLGGLVAVLIVLGL----- 295
D 209 GAWLAVLIVGVMLAILNRQW-----KLIG-VGAILLAIIGALVITQHN 251
OY 296 -BPLRVRLSLFVGRSDS--NNFRINWMLAVLQIDRPWIGPGTAFNLVPL 349
D 252 KRPD-----EHLTYLQQTDSYRTNCTGTAMI-----LIQENPIGYGVNDVIGYINK 304
OY 350 -----YQOARF-TALSAYSVPLEVAEGGLLGTAFAML-----LTVATAVQVSRRLRD 400
D 305 RVVDYPTWTFEKESIGPHNTILYIWFSAHGLASLVLYGAIIRRTASTLRKV-----E 359
OY 401 RNPQAFWIMASLAGLMIGHLPDVLV 427
D 360 ISPYNAHLILFLSFVGEYIVRGNFQV 386

```

## RESULT 4

```

ID NUON_ECOLI STANDARD: PRT: 425 AA.
AC P33608; P78281;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE NADH DEHYDROGENASE I CHAIN N (EC 1.6.5.3) (NADH-UBIQUINONE
DE OXIDOREDUCTASE CHAIN N (NADH14).
GN NUON OR B2276 OR Z3534 OR ECS3160.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / AN387;
RX MEDLINE=93389724; PubMed=7690854;
RA Weiner U., Geier S., Plock A., Friedrich T., Leif H., Weiss H.;
RT "The gene locus of the proton-translocating NADH: ubiquinone
RT oxidoreductase in Escherichia coli. Organization of the 14 genes and
RT relationship between the derived proteins and subunits of
RT mitochondrial complex I.";
RL J. Mol. Biol. 233:109-122(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97342617; PubMed=9278503;
RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubram S.,

```

```

FT   CONFLICT 208      217      AIRRVLAIIA -> QVRRCAGYR (IN REF. 1).
SO   SEQUENCE 425 AA; 45650 MM; 981BD4DFEED08B1F CRC64;

Query Match          5.0%, Score 122, DB 1; Length 425;
Best Local Similarity 23.3%, Pred. No. 0.44;
Matches 82; Conservative 48; Mismatches 114; Indels 108; Gaps 19;

OY  27  FGSLLRRMRSSQLLWSEALGGFL---AAVY---GSAPPV-----PSSALG 67
      |  :  |  :  :  :  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB   87  FPROKRSLEMSIKYTTLSIAASSFFLLFGMALVYADSGSLSVALLGRNLGDMLEPLLALG 146
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY  68  LGLAIAAAYWALLSLTDIDLROATPIH-WLWLYWGYDA-LATGISPVRAALVGLA-KL 124
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB   147  FGLMIVGLGPKL-----SLVFHLMPDPDYOGAPAEVSTPLATASKAIFGVVRL 197
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY  125  TLV-----LWVLAALARLNR-----PRLKSLFVVVIT 154
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB   198  FLVAPWGDSBAIRVLAIIAEFASIIIEGMLMALSGTNIKRLGYSISHLGVLVLAIILO 257
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY  155  SLFVSVYGLNOMI-----YGEVELATWDRNSVADFTSRVSYLG---NPNLAA 201
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB   258  TGEISMENAVGYIAGLIFSLSLGAFCFVSILMSPYRGPDAD---SLRSYGLFMHPRIIA 314
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY  202  LYVETTAFFSAAIAGVRCRGLPKLLAIATGASSICLLITTSRCGRNLGFPAAMFFWALLGL 261
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB   315  -VMTVMMLSLAGIPMTLGIIGFIFYVLAVGVAHL-----WML--GGAVVVSAIGL 362
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY  262  YWF-----QPLRPAPMRRLFPVVLGG---LVAVILLVVLGLER 297
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB   363  YYIARVAVSLYLHAPRGPRDAP-SNMGYSA--GGIVVLLSALLVVLGVMP 411
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT  5
YK26_PSEAE          STANDARD:          PRT:   333 AA.
ID   YK26_PSEAE          STANDARD:          PRT:   333 AA.
AC   P39879; O91289;
DT   01-FEB-1995 (Rel. 31, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN PA2026.
GN   PA2026.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID:287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-AFCC 15692 / PA01;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA   Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA   Bader L.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yan Y.,
RA   Grody R.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
RN   [2]
RP   SEQUENCE OF 281-333 FROM N.A.
RC   STRAIN-PA08;
RX   MEDLINE=91194546; PubMed=1849605;
RA   Perry A.C.F., Bhairain N.N., Brown N.L., Rouch D.A.;
RT   "Molecular characterization of the gor gene encoding glutathione
RT   reductase from Pseudomonas aeruginosa: determinants of substrate
RT   specificity among pyridine nucleotide-disulphide oxidoreductases.";
RL   Mol. Microbiol. 5:163-171(1991).
CC   -1 SIMILARITY: TO E.COLI YEFH.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC   the European Bioinformatics Institute. There are no restrictions on its

```



```

Db 110 YWLLSLGLVLLASFE-----TESGAPNVGTLVAPLTARIYT-----PGIGLD 153
QY 199 LAATVPTTAFSAALIGWRGMLPKLLAIAATGASSLCILITTSRGWLGFPVAMIFWAL 258
Db 154 LAA-----LATFLFSLSTYLTGTINFLVIAAMRAPGI-----GM--FKMPFTWSI 197
QY 259 IGLWFQPRLPAPWRMRMLFPVVLGVLAVLLVAVLGLE----- 296
Db 198 L-----FTYIIMLMAPFPLMVGALLLDNRIGTFEFLNPAGCALMDHLEWFGHP 249
QY 297 -----PLRVRLSTFVGREDSSNNFRINWLAVALQIQRPMGLG--IGPGNTAF 343
Db 250 EYVILFPALGAMADVISTFSGKPIYAKRYILTAFL--IATYISFVVMHMFITGTNIYT 308
QY 344 NIVVPL-----YQQA-----RPTALSAVSVPLEVAVEGGLGLTAFAMILLVAV 388
Db 309 RLEYSITTTLLISIPPEMAWMSFTFLYKRGLYTVPMLFPAV-GALLN-----FTIGGS 360
QY 389 TAVROYSRLRDRNPQ-AFWLMASLAG-LAGMLGHGLFDTVLVYRPEASTLWV 438
Db 361 TGV-YIGSLAIDRGFGRTYVVAHFHYLLVGVTLGLGLAGLY-----WM 404

RESULT 7
NU4M_STRPU
ID NU4M_STRPU STANDARD; PRT; 463 AA.
AC P15551;
DT 01-APR-1990 (rel. 14, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3).
GN ND4.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Mitochondrion.
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OC NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=89011951; PubMed=3172215;
RA Jacobs H.T., Elliott D.J., Malt V.B., Farquharson A.;
RT "Nucleotide sequence and gene organization of sea urchin
RT mitochondrial DNA.";
RL J. Mol. Biol. 202:185-217(1988).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X12631; CAA1159.2; -.
CC PIR: S01508; S01508.
CC InterPro: IPR003918; NADHUB_oxdrctset4.
CC InterPro: IPR001750; Oxidored_g1.
CC InterPro: IPR000260; Oxidored_g5_N.
CC Pfam: PF00361; oxidored_g1; 1.
CC Pfam: PF01059; oxidored_g5_N; 1.
CC Oxidoreductase; NAD; ubiquinone; Mitochondrion.
CC SROUNCE 463 AA; 50507 MW; 5C8E73FE7D5CAAF0 CNC64;

Query Match 4.9%; Score 119; DB 1; Length 463;
Best local similarity 18.8%; Pred. No. 0.73;
Matches 90; Conservative 52; Mismatches 132; Indels 204; Gaps 17;

QY 61 VPSAAGLGIAATAYWALLSTDDLRQATPIHW-----LVLLY 100
Db 19 IPSNKLWAGATFOSALLSLSLIVLN-----NHWTAWHNLSTLASDGTISAPLLITLSC 72

```

```

QY 101 MGVDAIATGSLSPVRAALVGL-----AKITLYLLVFALAAVLRNPLRSLSFVVV 152
Db 73 W-----LAPIALIASKQOLNNSDLSGRVIFIMITYTGAALITPSSLEILEIYIV 124
QY 153 ITSLSFVSYYGLNQWITYGVEELATVWDRNSVADFTSRVSYLGNPNLLAAYVFTTFESA 212
Db 125 EFTTLPTLLILYIPWCAQMERCOA-----GLYFMFTLF-- 157
QY 213 ALGVWRGMLPKLLAIAATGASSLCILITYSRGWL--GFVAMIFVWALLGLYFQPRLPA 270
Db 158 -----GSLPLIALIAYISSSLSPNVNLLMANDGSIESLTMWALSTNCFENMLPY 211
QY 271 -PWRMRL-----FPVVLGLVAVLLVAVLGLEPLRVRLSTFVGREDSSNNFRINWMLA 323
Db 212 YGFHMLPKRAHEAPVAVGSMILAILKIGGYGLMR----- 247
QY 324 VIQMIQDRPMWLGIGGNTAFNLVYPLVQQAFTLSAJSVPLEVAVGGILGTAF-AML 382
Db 248 -----LALPFTISMNALS-----LALIVFTWIG 271
QY 383 LVTAVTAVROYSRLRDRNPQAFWLMASLA-----GLAG--M 418
Db 272 ALITSYICVRQID-----LKAALITYSSGVHMSIYAAAISETSMGNALMLM 319
QY 419 LGHGLFDLYL-----RPEA-----STLMWLCIGALASFQPOPS 453
Db 320 VAHGLVSSALFPLANTYERSGTRLATRGLKLLPLSTLMWLMCAANILGPPSPN 377

RESULT 8
Y608_HAEIN
ID Y608_HAEIN STANDARD; PRT; 461 AA.
AC 057486;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN H10608.
GN H10608.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=RD / KM20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -I- SIMILARITY: BELONGS TO THE NAD(P)/PHO8 FAMILY OF TRANSPORTERS.
CC NADC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32743; AAC22267.1; -.
CC TIGR: H10608; -.

```

DR InterPro: IPR001898; Na.sulph. symp.  
 DR Pfam: PF00939; Na.sulph. symp. 3.  
 DR PROSITE: PS01271; Na.SULFATP; 1.  
 KW Hypothetical protein; Transmembrane; Transport; Complete proteome.  
 FT TRANSMEM 13  
 FT TRANSMEM 33 POTENTIAL.  
 FT TRANSMEM 54 74 POTENTIAL.  
 FT TRANSMEM 81 101 POTENTIAL.  
 FT TRANSMEM 120 140 POTENTIAL.  
 FT TRANSMEM 170 190 POTENTIAL.  
 FT TRANSMEM 211 221 POTENTIAL.  
 FT TRANSMEM 256 276 POTENTIAL.  
 FT TRANSMEM 286 306 POTENTIAL.  
 FT TRANSMEM 314 334 POTENTIAL.  
 FT TRANSMEM 349 369 POTENTIAL.  
 FT TRANSMEM 377 397 POTENTIAL.  
 FT TRANSMEM 399 419 POTENTIAL.  
 FT TRANSMEM 439 459 POTENTIAL.  
 SQ SEQUENCE 461 AA; 49761 MW; B5E6F6965B38EF06 CRC64;

Query Match 4.9%; Score 118.5; DB 1; Length 461;  
 Best Local Similarity 23.2%; Pred. No. 0.78;  
 Matches 110; Conservative 64; Mismatches 165; Indels 135; Gaps 27;

QY 50 LLAIVYGSAPFYPPSSALGIGLAA-IAAYWALLSTFDIDLRQATPIHMLVLYGWYDALAT 108  
 DB 22 LFEVLINLVLPFEPKANSGLALAFIAVLW-----LSEALHTITALL--VPLAV 69  
 QY 109 GLSPARA-AALVGLAKLITYLYL--PALAA-----RVLRNRL--RSLLFVYVIT 154  
 DB 70 ALGLVSTKQALVGFADPTLFFFGFSLATLHIOKIDILINKIMALRGMFLTAIVY- 128  
 QY 155 SLEFVSYGGLNOMIVGEELA-----TWVDNSVADFSRYSYLGNNMLAIVY 204  
 DB 129 -LFILTAFLSMNSMTATAPAMMLPLAMGLISQIDREK--DHNTYFVLG----- 175  
 QY 205 PPTAASAA--AIGVWRGMLPKLLAIAATGASSLCLITYSRGCGVGFAMIFVWL-LGL 261  
 DB 176 --IAVSASIGMGTLVGSPPN--AIVASN-----LNLFSDDLWGLPIMILLPLMIGI 226  
 QY 264 YW--FQPRLPAPWRM-----LFPV-----VLGIVAVLVAVIGLE--- 296  
 DB 227 LVTIKPKHLNFEOTFENIENPMRILTLFIFPVIALTWFSKINPISGLIDLOKNI 286  
 QY 297 ---PLRVLSIFVGRDESSNNFRINWMLAVLQMIODRPLGTI---GPGNTAFNLVYPL 349  
 DB 287 ASFSDIVALLAIY-----ICSTGVASWKQIQSNTDMGVLMLEGGGLT---LSAV 333  
 QY 330 YQQAFFTALSAVSPLLEVAVEGG---LGLTAFAMLLVTVAVTAVROVSRLEDRNPQAF 406  
 DB 334 LKDSGASKILADSIYF--MIDGQHYLLGLLVAFAFIPLTEPTS-----NTASAA 381  
 QY 407 WLMAVLGAGLACGLGFLDYIVYRPEASTLWMLCIGALASFPQOPSPQLPPEA 460  
 DB 382 LVIPLIFISTAGSLG-----MPEIGLALITIGASCAPMLPVAT---PPNA 423

RESULT 9  
 ID LAT2\_HUMAN STANDARD: PRT; 535 AA.  
 AC Q9UHI5; Q9Y445; Q9UKQ6; Q9UKQ7; Q9UKQ8;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE LARGE NEUTRAL AMINO ACIDS TRANSPORTER SMALL SUBUNIT 2 (L-TYPE AMINO ACID TRANSPORTER 2).  
 GN SLG7A8 OR LAT2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE-20044753; PubMed-10574970;  
 RA Rossier G., Meier C., Bauch C., Summa V., Sordal B., Verrey F.,  
 RA Kuhn L.C.;  
 RT "LAT2, a new basolateral 4F2hc/CD98-associated amino acid transporter  
 RT of kidney and intestine."  
 RL J. Biol. Chem. 274:34948-34954(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE-99178268; PubMed-10080183;  
 RA Borsani G., Bassi M.T., Sperandeo M.P., De Grandi A., Buoninconti A.,  
 RA Riboni M., Manzoni M., Incerli B., Pepe A., Andria G., Ballabio A.,  
 RA Sebastio G.;  
 RT "SLC7A7, encoding a putative permease-related protein, is mutated in  
 RT patients with lysinuric protein intolerance."  
 RL Nat. Genet. 21:297-301(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99321901; PubMed-10391915;  
 RA Plineda M., Fernandez E., Torrents D., Estevez R., Lopez C., Camps M.,  
 RA Lloberas J., Zorzano A., Palacin M.;  
 RT "Identification of a membrane protein, LAT-2, that co-expresses with  
 RT 4F2 heavy chain, an L-type amino acid transport activity with broad  
 RT specificity for small and large zwitterionic amino acids."  
 RL J. Biol. Chem. 274:19738-19744(1999).  
 CC -1- FUNCTION: SODIUM-INDEPENDENT, HIGH-AFFINITY TRANSPORT OF LARGE  
 CC NEUTRAL AMINO ACIDS. HAS HIGHER AFFINITY FOR L-PHENYLALANINE THAT  
 CC LAT1. L-ALANINE IS TRANSPORTED AT PHYSIOLOGICAL CONCENTRATIONS.  
 CC PLAYS A ROLE IN BASOLATERAL (RE)ABSORPTION OF NEUTRAL AMINO ACIDS.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER WITH THE AMINO ACID  
 CC TRANSPORT PROTEIN SLC3A2/4F2HC.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLY).  
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN KIDNEY AND INTESTINE.  
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MANY AMINO ACID  
 CC PERMEASES. GLYCOPROTEIN-ASSOCIATED AMINO ACID TRANSPORTERS (GPAT)  
 CC SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; AF171669; AAF20381.1; -  
 CC EMBL; Y18483; CAB40137.1; -  
 CC EMBL; AF135830; AAF05697.1; -  
 CC EMBL; AF135829; AAF05696.1; -  
 CC EMBL; AF135828; AAF05695.1; -  
 CC MIM; 604235; -  
 DR InterPro: IPR002293; AA\_rel\_permease\_1.  
 DR InterPro: IPR002027; amino\_acid\_permease.  
 DR Pfam: PF00324; aa\_permeases; 1.  
 KW Transport; Amino-acid transport; Transmembrane.  
 FT TRANSMEM 40  
 FT TRANSMEM 60 POTENTIAL.  
 FT TRANSMEM 72 92 POTENTIAL.  
 FT TRANSMEM 113 133 POTENTIAL.  
 FT TRANSMEM 155 175 POTENTIAL.  
 FT TRANSMEM 189 209 POTENTIAL.  
 FT TRANSMEM 231 251 POTENTIAL.  
 FT TRANSMEM 268 288 POTENTIAL.  
 FT TRANSMEM 310 330 POTENTIAL.  
 FT TRANSMEM 362 382 POTENTIAL.  
 FT TRANSMEM 388 408 POTENTIAL.  
 FT TRANSMEM 424 444 POTENTIAL.  
 FT TRANSMEM 447 467 POTENTIAL.  
 FT TRANSMEM 475 495 POTENTIAL.  
 FT TRANSMEM 495 515 POTENTIAL.  
 FT TRANSMEM 515 535 POTENTIAL.  
 FT TRANSMEM 535 555 POTENTIAL.  
 FT TRANSMEM 555 575 POTENTIAL.  
 FT TRANSMEM 575 595 POTENTIAL.  
 FT TRANSMEM 595 615 POTENTIAL.  
 FT TRANSMEM 615 635 POTENTIAL.  
 FT TRANSMEM 635 655 POTENTIAL.  
 FT TRANSMEM 655 675 POTENTIAL.  
 FT TRANSMEM 675 695 POTENTIAL.  
 FT TRANSMEM 695 715 POTENTIAL.  
 FT TRANSMEM 715 735 POTENTIAL.  
 FT TRANSMEM 735 755 POTENTIAL.  
 FT TRANSMEM 755 775 POTENTIAL.  
 FT TRANSMEM 775 795 POTENTIAL.  
 FT TRANSMEM 795 815 POTENTIAL.  
 FT TRANSMEM 815 835 POTENTIAL.  
 FT TRANSMEM 835 855 POTENTIAL.  
 FT TRANSMEM 855 875 POTENTIAL.  
 FT TRANSMEM 875 895 POTENTIAL.  
 FT TRANSMEM 895 915 POTENTIAL.  
 FT TRANSMEM 915 935 POTENTIAL.  
 FT TRANSMEM 935 955 POTENTIAL.  
 FT TRANSMEM 955 975 POTENTIAL.  
 FT TRANSMEM 975 995 POTENTIAL.  
 FT TRANSMEM 995 1015 POTENTIAL.  
 FT TRANSMEM 1015 1035 POTENTIAL.  
 FT TRANSMEM 1035 1055 POTENTIAL.  
 FT TRANSMEM 1055 1075 POTENTIAL.  
 FT TRANSMEM 1075 1095 POTENTIAL.  
 FT TRANSMEM 1095 1115 POTENTIAL.  
 FT TRANSMEM 1115 1135 POTENTIAL.  
 FT TRANSMEM 1135 1155 POTENTIAL.  
 FT TRANSMEM 1155 1175 POTENTIAL.  
 FT TRANSMEM 1175 1195 POTENTIAL.  
 FT TRANSMEM 1195 1215 POTENTIAL.  
 FT TRANSMEM 1215 1235 POTENTIAL.  
 FT TRANSMEM 1235 1255 POTENTIAL.  
 FT TRANSMEM 1255 1275 POTENTIAL.  
 FT TRANSMEM 1275 1295 POTENTIAL.  
 FT TRANSMEM 1295 1315 POTENTIAL.  
 FT TRANSMEM 1315 1335 POTENTIAL.  
 FT TRANSMEM 1335 1355 POTENTIAL.  
 FT TRANSMEM 1355 1375 POTENTIAL.  
 FT TRANSMEM 1375 1395 POTENTIAL.  
 FT TRANSMEM 1395 1415 POTENTIAL.  
 FT TRANSMEM 1415 1435 POTENTIAL.  
 FT TRANSMEM 1435 1455 POTENTIAL.  
 FT TRANSMEM 1455 1475 POTENTIAL.  
 FT TRANSMEM 1475 1495 POTENTIAL.  
 FT TRANSMEM 1495 1515 POTENTIAL.  
 FT TRANSMEM 1515 1535 POTENTIAL.  
 FT TRANSMEM 1535 1555 POTENTIAL.  
 FT TRANSMEM 1555 1575 POTENTIAL.  
 FT TRANSMEM 1575 1595 POTENTIAL.  
 FT TRANSMEM 1595 1615 POTENTIAL.  
 FT TRANSMEM 1615 1635 POTENTIAL.  
 FT TRANSMEM 1635 1655 POTENTIAL.  
 FT TRANSMEM 1655 1675 POTENTIAL.  
 FT TRANSMEM 1675 1695 POTENTIAL.  
 FT TRANSMEM 1695 1715 POTENTIAL.  
 FT TRANSMEM 1715 1735 POTENTIAL.  
 FT TRANSMEM 1735 1755 POTENTIAL.  
 FT TRANSMEM 1755 1775 POTENTIAL.  
 FT TRANSMEM 1775 1795 POTENTIAL.  
 FT TRANSMEM 1795 1815 POTENTIAL.  
 FT TRANSMEM 1815 1835 POTENTIAL.  
 FT TRANSMEM 1835 1855 POTENTIAL.  
 FT TRANSMEM 1855 1875 POTENTIAL.  
 FT TRANSMEM 1875 1895 POTENTIAL.  
 FT TRANSMEM 1895 1915 POTENTIAL.  
 FT TRANSMEM 1915 1935 POTENTIAL.  
 FT TRANSMEM 1935 1955 POTENTIAL.  
 FT TRANSMEM 1955 1975 POTENTIAL.  
 FT TRANSMEM 1975 1995 POTENTIAL.  
 FT TRANSMEM 1995 2015 POTENTIAL.  
 FT TRANSMEM 2015 2035 POTENTIAL.  
 FT TRANSMEM 2035 2055 POTENTIAL.  
 FT TRANSMEM 2055 2075 POTENTIAL.  
 FT TRANSMEM 2075 2095 POTENTIAL.  
 FT TRANSMEM 2095 2115 POTENTIAL.  
 FT TRANSMEM 2115 2135 POTENTIAL.  
 FT TRANSMEM 2135 2155 POTENTIAL.  
 FT TRANSMEM 2155 2175 POTENTIAL.  
 FT TRANSMEM 2175 2195 POTENTIAL.  
 FT TRANSMEM 2195 2215 POTENTIAL.  
 FT TRANSMEM 2215 2235 POTENTIAL.  
 FT TRANSMEM 2235 2255 POTENTIAL.  
 FT TRANSMEM 2255 2275 POTENTIAL.  
 FT TRANSMEM 2275 2295 POTENTIAL.  
 FT TRANSMEM 2295 2315 POTENTIAL.  
 FT TRANSMEM 2315 2335 POTENTIAL.  
 FT TRANSMEM 2335 2355 POTENTIAL.  
 FT TRANSMEM 2355 2375 POTENTIAL.  
 FT TRANSMEM 2375 2395 POTENTIAL.  
 FT TRANSMEM 2395 2415 POTENTIAL.  
 FT TRANSMEM 2415 2435 POTENTIAL.  
 FT TRANSMEM 2435 2455 POTENTIAL.  
 FT TRANSMEM 2455 2475 POTENTIAL.  
 FT TRANSMEM 2475 2495 POTENTIAL.  
 FT TRANSMEM 2495 2515 POTENTIAL.  
 FT TRANSMEM 2515 2535 POTENTIAL.  
 FT TRANSMEM 2535 2555 POTENTIAL.  
 FT TRANSMEM 2555 2575 POTENTIAL.  
 FT TRANSMEM 2575 2595 POTENTIAL.  
 FT TRANSMEM 2595 2615 POTENTIAL.  
 FT TRANSMEM 2615 2635 POTENTIAL.  
 FT TRANSMEM 2635 2655 POTENTIAL.  
 FT TRANSMEM 2655 2675 POTENTIAL.  
 FT TRANSMEM 2675 2695 POTENTIAL.  
 FT TRANSMEM 2695 2715 POTENTIAL.  
 FT TRANSMEM 2715 2735 POTENTIAL.  
 FT TRANSMEM 2735 2755 POTENTIAL.  
 FT TRANSMEM 2755 2775 POTENTIAL.  
 FT TRANSMEM 2775 2795 POTENTIAL.  
 FT TRANSMEM 2795 2815 POTENTIAL.  
 FT TRANSMEM 2815 2835 POTENTIAL.  
 FT TRANSMEM 2835 2855 POTENTIAL.  
 FT TRANSMEM 2855 2875 POTENTIAL.  
 FT TRANSMEM 2875 2895 POTENTIAL.  
 FT TRANSMEM 2895 2915 POTENTIAL.  
 FT TRANSMEM 2915 2935 POTENTIAL.  
 FT TRANSMEM 2935 2955 POTENTIAL.  
 FT TRANSMEM 2955 2975 POTENTIAL.  
 FT TRANSMEM 2975 2995 POTENTIAL.  
 FT TRANSMEM 2995 3015 POTENTIAL.  
 FT TRANSMEM 3015 3035 POTENTIAL.  
 FT TRANSMEM 3035 3055 POTENTIAL.  
 FT TRANSMEM 3055 3075 POTENTIAL.  
 FT TRANSMEM 3075 3095 POTENTIAL.  
 FT TRANSMEM 3095 3115 POTENTIAL.  
 FT TRANSMEM 3115 3135 POTENTIAL.  
 FT TRANSMEM 3135 3155 POTENTIAL.  
 FT TRANSMEM 3155 3175 POTENTIAL.  
 FT TRANSMEM 3175 3195 POTENTIAL.  
 FT TRANSMEM 3195 3215 POTENTIAL.  
 FT TRANSMEM 3215 3235 POTENTIAL.  
 FT TRANSMEM 3235 3255 POTENTIAL.  
 FT TRANSMEM 3255 3275 POTENTIAL.  
 FT TRANSMEM 3275 3295 POTENTIAL.  
 FT TRANSMEM 3295 3315 POTENTIAL.  
 FT TRANSMEM 3315 3335 POTENTIAL.  
 FT TRANSMEM 3335 3355 POTENTIAL.  
 FT TRANSMEM 3355 3375 POTENTIAL.  
 FT TRANSMEM 3375 3395 POTENTIAL.  
 FT TRANSMEM 3395 3415 POTENTIAL.  
 FT TRANSMEM 3415 3435 POTENTIAL.  
 FT TRANSMEM 3435 3455 POTENTIAL.  
 FT TRANSMEM 3455 3475 POTENTIAL.  
 FT TRANSMEM 3475 3495 POTENTIAL.  
 FT TRANSMEM 3495 3515 POTENTIAL.  
 FT TRANSMEM 3515 3535 POTENTIAL.  
 FT TRANSMEM 3535 3555 POTENTIAL.  
 FT TRANSMEM 3555 3575 POTENTIAL.  
 FT TRANSMEM 3575 3595 POTENTIAL.  
 FT TRANSMEM 3595 3615 POTENTIAL.  
 FT TRANSMEM 3615 3635 POTENTIAL.  
 FT TRANSMEM 3635 3655 POTENTIAL.  
 FT TRANSMEM 3655 3675 POTENTIAL.  
 FT TRANSMEM 3675 3695 POTENTIAL.  
 FT TRANSMEM 3695 3715 POTENTIAL.  
 FT TRANSMEM 3715 3735 POTENTIAL.  
 FT TRANSMEM 3735 3755 POTENTIAL.  
 FT TRANSMEM 3755 3775 POTENTIAL.  
 FT TRANSMEM 3775 3795 POTENTIAL.  
 FT TRANSMEM 3795 3815 POTENTIAL.  
 FT TRANSMEM 3815 3835 POTENTIAL.  
 FT TRANSMEM 3835 3855 POTENTIAL.  
 FT TRANSMEM 3855 3875 POTENTIAL.  
 FT TRANSMEM 3875 3895 POTENTIAL.  
 FT TRANSMEM 3895 3915 POTENTIAL.  
 FT TRANSMEM 3915 3935 POTENTIAL.  
 FT TRANSMEM 3935 3955 POTENTIAL.  
 FT TRANSMEM 3955 3975 POTENTIAL.  
 FT TRANSMEM 3975 3995 POTENTIAL.  
 FT TRANSMEM 3995 4015 POTENTIAL.  
 FT TRANSMEM 4015 4035 POTENTIAL.  
 FT TRANSMEM 4035 4055 POTENTIAL.  
 FT TRANSMEM 4055 4075 POTENTIAL.  
 FT TRANSMEM 4075 4095 POTENTIAL.  
 FT TRANSMEM 4095 4115 POTENTIAL.  
 FT TRANSMEM 4115 4135 POTENTIAL.  
 FT TRANSMEM 4135 4155 POTENTIAL.  
 FT TRANSMEM 4155 4175 POTENTIAL.  
 FT TRANSMEM 4175 4195 POTENTIAL.  
 FT TRANSMEM 4195 4215 POTENTIAL.  
 FT TRANSMEM 4215 4235 POTENTIAL.  
 FT TRANSMEM 4235 4255 POTENTIAL.  
 FT TRANSMEM 4255 4275 POTENTIAL.  
 FT TRANSMEM 4275 4295 POTENTIAL.  
 FT TRANSMEM 4295 4315 POTENTIAL.  
 FT TRANSMEM 4315 4335 POTENTIAL.  
 FT TRANSMEM 4335 4355 POTENTIAL.  
 FT TRANSMEM 4355 4375 POTENTIAL.  
 FT TRANSMEM 4375 4395 POTENTIAL.  
 FT TRANSMEM 4395 4415 POTENTIAL.  
 FT TRANSMEM 4415 4435 POTENTIAL.  
 FT TRANSMEM 4435 4455 POTENTIAL.  
 FT TRANSMEM 4455 4475 POTENTIAL.  
 FT TRANSMEM 4475 4495 POTENTIAL.  
 FT TRANSMEM 4495 4515 POTENTIAL.  
 FT TRANSMEM 4515 4535 POTENTIAL.  
 FT TRANSMEM 4535 4555 POTENTIAL.  
 FT TRANSMEM 4555 4575 POTENTIAL.  
 FT TRANSMEM 4575 4595 POTENTIAL.  
 FT TRANSMEM 4595 4615 POTENTIAL.  
 FT TRANSMEM 4615 4635 POTENTIAL.  
 FT TRANSMEM 4635 4655 POTENTIAL.  
 FT TRANSMEM 4655 4675 POTENTIAL.  
 FT TRANSMEM 4675 4695 POTENTIAL.  
 FT TRANSMEM 4695 4715 POTENTIAL.  
 FT TRANSMEM 4715 4735 POTENTIAL.  
 FT TRANSMEM 4735 4755 POTENTIAL.  
 FT TRANSMEM 4755 4775 POTENTIAL.  
 FT TRANSMEM 4775 4795 POTENTIAL.  
 FT TRANSMEM 4795 4815 POTENTIAL.  
 FT TRANSMEM 4815 4835 POTENTIAL.  
 FT TRANSMEM 4835 4855 POTENTIAL.  
 FT TRANSMEM 4855 4875 POTENTIAL.  
 FT TRANSMEM 4875 4895 POTENTIAL.  
 FT TRANSMEM 4895 4915 POTENTIAL.  
 FT TRANSMEM 4915 4935 POTENTIAL.  
 FT TRANSMEM 4935 4955 POTENTIAL.  
 FT TRANSMEM 4955 4975 POTENTIAL.  
 FT TRANSMEM 4975 4995 POTENTIAL.  
 FT TRANSMEM 4995 5015 POTENTIAL.  
 FT TRANSMEM 5015 5035 POTENTIAL.  
 FT TRANSMEM 5035 5055 POTENTIAL.  
 FT TRANSMEM 5055 5075 POTENTIAL.  
 FT TRANSMEM 5075 5095 POTENTIAL.  
 FT TRANSMEM 5095 5115 POTENTIAL.  
 FT TRANSMEM 5115 5135 POTENTIAL.  
 FT TRANSMEM 5135 5155 POTENTIAL.  
 FT TRANSMEM 5155 5175 POTENTIAL.  
 FT TRANSMEM 5175 5195 POTENTIAL.  
 FT TRANSMEM 5195 5215 POTENTIAL.  
 FT TRANSMEM 5215 5235 POTENTIAL.  
 FT TRANSMEM 5235 5255 POTENTIAL.  
 FT TRANSMEM 5255 5275 POTENTIAL.  
 FT TRANSMEM 5275 5295 POTENTIAL.  
 FT TRANSMEM 5295 5315 POTENTIAL.  
 FT TRANSMEM 5315 5335 POTENTIAL.  
 FT TRANSMEM 5335 5355 POTENTIAL.  
 FT TRANSMEM 5355 5375 POTENTIAL.  
 FT TRANSMEM 5375 5395 POTENTIAL.  
 FT TRANSMEM 5395 5415 POTENTIAL.  
 FT TRANSMEM 5415 5435 POTENTIAL.  
 FT TRANSMEM 5435 5455 POTENTIAL.  
 FT TRANSMEM 5455 5475 POTENTIAL.  
 FT TRANSMEM 5475 5495 POTENTIAL.  
 FT TRANSMEM 5495 5515 POTENTIAL.  
 FT TRANSMEM 5515 5535 POTENTIAL.  
 FT TRANSMEM 5535 5555 POTENTIAL.  
 FT TRANSMEM 5555 5575 POTENTIAL.  
 FT TRANSMEM 5575 5595 POTENTIAL.  
 FT TRANSMEM 5595 5615 POTENTIAL.  
 FT TRANSMEM 5615 5635 POTENTIAL.  
 FT TRANSMEM 5635 5655 POTENTIAL.  
 FT TRANSMEM 5655 5675 POTENTIAL.  
 FT TRANSMEM 5675 5695 POTENTIAL.  
 FT TRANSMEM 5695 5715 POTENTIAL.  
 FT TRANSMEM 5715 5735 POTENTIAL.  
 FT TRANSMEM 5735 5755 POTENTIAL.  
 FT TRANSMEM 5755 5775 POTENTIAL.  
 FT TRANSMEM 5775 5795 POTENTIAL.  
 FT TRANSMEM 5795 5815 POTENTIAL.  
 FT TRANSMEM 5815 5835 POTENTIAL.  
 FT TRANSMEM 5835 5855 POTENTIAL.  
 FT TRANSMEM 5855 5875 POTENTIAL.  
 FT TRANSMEM 5875 5895 POTENTIAL.  
 FT TRANSMEM 5895 5915 POTENTIAL.  
 FT TRANSMEM 5915 5935 POTENTIAL.  
 FT TRANSMEM 5935 5955 POTENTIAL.  
 FT TRANSMEM 5955 5975 POTENTIAL.  
 FT TRANSMEM 5975 5995 POTENTIAL.  
 FT TRANSMEM 5995 6015 POTENTIAL.  
 FT TRANSMEM 6015 6035 POTENTIAL.  
 FT TRANSMEM 6035 6055 POTENTIAL.  
 FT TRANSMEM 6055 6075 POTENTIAL.  
 FT TRANSMEM 6075 6095 POTENTIAL.  
 FT TRANSMEM 6095 6115 POTENTIAL.  
 FT TRANSMEM 6115 6135 POTENTIAL.  
 FT TRANSMEM 6135 6155 POTENTIAL.  
 FT TRANSMEM 6155 6175 POTENTIAL.  
 FT TRANSMEM 6175 6195 POTENTIAL.  
 FT TRANSMEM 6195 6215 POTENTIAL.  
 FT TRANSMEM 6215 6235 POTENTIAL.  
 FT TRANSMEM 6235 6255 POTENTIAL.  
 FT TRANSMEM 6255 6275 POTENTIAL.  
 FT TRANSMEM 6275 6295 POTENTIAL.  
 FT TRANSMEM 6295 6315 POTENTIAL.  
 FT TRANSMEM 6315 6335 POTENTIAL.  
 FT TRANSMEM 6335 6355 POTENTIAL.  
 FT TRANSMEM 6355 6375 POTENTIAL.  
 FT TRANSMEM 6375 6395 POTENTIAL.  
 FT TRANSMEM 6395 6415 POTENTIAL.  
 FT TRANSMEM 6415 6435 POTENTIAL.  
 FT TRANSMEM 6435 6455 POTENTIAL.  
 FT TRANSMEM 6455 6475 POTENTIAL.  
 FT TRANSMEM 6475 6495 POTENTIAL.  
 FT TRANSMEM 6495 6515 POTENTIAL.  
 FT TRANSMEM 6515 6535 POTENTIAL.  
 FT TRANSMEM 6535 6555 POTENTIAL.  
 FT TRANSMEM 6555 6575 POTENTIAL.  
 FT TRANSMEM 6575 6595 POTENTIAL.  
 FT TRANSMEM 6595 6615 POTENTIAL.  
 FT TRANSMEM 6615 6635 POTENTIAL.  
 FT TRANSMEM 6635 6655 POTENTIAL.  
 FT TRANSMEM 6655 6675 POTENTIAL.  
 FT TRANSMEM 6675 6695 POTENTIAL.  
 FT TRANSMEM 6695 6715 POTENTIAL.  
 FT TRANSMEM 6715 6735 POTENTIAL.  
 FT TRANSMEM 6735 6755 POTENTIAL.  
 FT TRANSMEM 6755 6775 POTENTIAL.  
 FT TRANSMEM 6775 6795 POTENTIAL.  
 FT TRANSMEM 6795 6815 POTENTIAL.  
 FT TRANSMEM 6815 6835 POTENTIAL.  
 FT TRANSMEM 6835 6855 POTENTIAL.  
 FT TRANSMEM 6855 6875 POTENTIAL.  
 FT TRANSMEM 6875 6895 POTENTIAL.  
 FT TRANSMEM 6895 6915 POTENTIAL.  
 FT TRANSMEM 6915 6935 POTENTIAL.  
 FT TRANSMEM 6935 6955 POTENTIAL.  
 FT TRANSMEM 6955 6975 POTENTIAL.  
 FT TRANSMEM 6975 6995 POTENTIAL.  
 FT TRANSMEM 6995 7015 POTENTIAL.  
 FT TRANSMEM 7015 7035 POTENTIAL.  
 FT TRANSMEM 7035 7055 POTENTIAL.  
 FT TRANSMEM 7055 7075 POTENTIAL.  
 FT TRANSMEM 7075 7095 POTENTIAL.  
 FT TRANSMEM 7095 7115 POTENTIAL.  
 FT TRANSMEM 7115 7135 POTENTIAL.  
 FT TRANSMEM 7135 7155 POTENTIAL.  
 FT TRANSMEM 7155 7175 POTENTIAL.  
 FT TRANSMEM 7175 7195 POTENTIAL.  
 FT TRANSMEM 7195 7215 POTENTIAL.  
 FT TRANSMEM 7215 7235 POTENTIAL.  
 FT TRANSMEM 7235 7255 POTENTIAL.  
 FT TRANSMEM 7255 7275 POTENTIAL.  
 FT TRANSMEM 7275 7295 POTENTIAL.  
 FT TRANSMEM 7295 7315 POTENTIAL.  
 FT TRANSMEM 7315 7335 POTENTIAL.  
 FT TRANSMEM 7335 7355 POTENTIAL.  
 FT TRANSMEM 7355 7375 POTENTIAL.  
 FT TRANSMEM 7375 7395 POTENTIAL.  
 FT TRANSMEM 7395 7415 POTENTIAL.  
 FT TRANSMEM 7415 7435 POTENTIAL.  
 FT TRANSMEM 7435 7455 POTENTIAL.  
 FT TRANSMEM 7455 7475 POTENTIAL.  
 FT TRANSMEM 7475 7495 POTENTIAL.  
 FT TRANSMEM 7495 7515 POTENTIAL.  
 FT TRANSMEM 7515 7535 POTENTIAL.  
 FT TRANSMEM 7535 7555 POTENTIAL.  
 FT TRANSMEM 7555 7575 POTENTIAL.  
 FT TRANSMEM 7575 7595 POTENTIAL.  
 FT TRANSMEM 7595 7615 POTENTIAL.  
 FT TRANSMEM 7615 7635 POTENTIAL.  
 FT TRANSMEM 7635 7655 POTENTIAL.  
 FT TRANSMEM 7655 7675 POTENTIAL.  
 FT TRANSMEM 7675 7695 POTENTIAL.  
 FT TRANSMEM 7695 7715 POTENTIAL.  
 FT TRANSMEM 7715 7735 POTENTIAL.  
 FT TRANSMEM 7735 7755 POTENTIAL.  
 FT TRANSMEM 7755 7775 POTENTIAL.  
 FT TRANSMEM 7775 7795 POTENTIAL.  
 FT TRANSMEM 7795 7815 POTENTIAL.  
 FT TRANSMEM 7815 7835 POTENTIAL.  
 FT TRANSMEM 7835 7855 POTENTIAL.  
 FT TRANSMEM 7855 7875 POTENTIAL.  
 FT TRANSMEM 7875 7895 POTENTIAL.  
 FT TRANSMEM 7895 7915 POTENTIAL.  
 FT TRANSMEM 7915 7935 POTENTIAL.  
 FT TRANSMEM 7935 7955 POTENTIAL.  
 FT TRANSMEM 7955 7975 POTENTIAL.  
 FT TRANSMEM 7975 7995 POTENTIAL.  
 FT TRANSMEM 7995 8015 POTENTIAL.  
 FT TRANSMEM 8015 8035 POTENTIAL.  
 FT TRANSMEM 8035 8055 POTENTIAL.  
 FT TRANSMEM 8055 8075 POTENTIAL.  
 FT TRANSMEM 8075 8095 POTENTIAL.  
 FT TRANSMEM 8095 8115 POTENTIAL.  
 FT TRANSMEM 8115 8135 POTENTIAL.  
 FT TRANSMEM 8135 8155 POTENTIAL.  
 FT TRANSMEM 8155 8175 POTENTIAL.  
 FT TRANSMEM 8175 8195 POTENTIAL.  
 FT TRANSMEM 8195 8215 POTENTIAL.  
 FT TRANSMEM 8215 8235 POTENTIAL.  
 FT TRANSMEM 8235 8255 POTENTIAL.  
 FT TRANSMEM 8255 8275 POTENTIAL.  
 FT TRANSMEM 8275 8295 POTENTIAL.  
 FT TRANSMEM 8295 8315 POTENTIAL.  
 FT TRANSMEM 8315 8335 POTENTIAL.  
 FT TRANSMEM 8335 8355 POTENTIAL.  
 FT TRANSMEM 8355 8375 POTENTIAL.  
 FT TRANSMEM 8375 8395 POTENTIAL.  
 FT TRANSMEM 8395 8415 POTENTIAL.  
 FT TRANSMEM 8415 8435 POTENTIAL.  
 FT TRANSMEM 8435 8455 POTENTIAL.  
 FT TRANSMEM 8455 8475 POTENTIAL.  
 FT TRANSMEM 8475 8495 POTENTIAL.  
 FT TRANSMEM 8495 8515 POTENTIAL.  
 FT TRANSMEM 8515 8535 POTENTIAL.  
 FT TRANSMEM 8535 8555 POTENTIAL.  
 FT TRANSMEM 8555 8575 POTENTIAL.  
 FT TRANSMEM 8575 8595 POTENTIAL.  
 FT TRANSMEM 8595 8615 POTENTIAL.  
 FT TRANSMEM 8615 8635 POTENTIAL.  
 FT TRANSMEM 8635 8655 POTENTIAL.  
 FT TRANSMEM 8655 8675 POTENTIAL.  
 FT TRANSMEM 8675 8695 POTENTIAL.  
 FT TRANSMEM 8695 8715 POTENTIAL.  
 FT TRANSMEM 8715 8735 POTENTIAL.  
 FT TRANSMEM 8735 8755 POTENTIAL.  
 FT TRANSMEM 8755 8775 POTENTIAL.  
 FT TRANSMEM 8775 8795 POTENTIAL.  
 FT TRANSMEM 8795 8815 POTENTIAL.  
 FT TRANSMEM 8815 8835 POTENTIAL.  
 FT TRANSMEM 8835 8855 POTENTIAL.  
 FT TRANSMEM 8855 8875 POTENTIAL.  
 FT TRANSMEM 8875 8895 POTENTIAL.  
 FT TRANSMEM 8895 8915 POTENTIAL.  
 FT TRANSMEM 8915 8935 POTENTIAL.  
 FT TRANSMEM 8935 8955 POTENTIAL.  
 FT TRANSMEM 8955 8975 POTENTIAL.  
 FT TRANSMEM 8975 8995 POTENTIAL.  
 FT TRANSMEM 8995 9015 POTENTIAL.  
 FT TRANSMEM 9015 9035 POTENTIAL.  
 FT TRANSMEM 9035 9055 POTENTIAL.  
 FT TRANSMEM 9055 9075 POTENTIAL.  
 FT TRANSMEM 9075 9095 POTENTIAL.  
 FT TRANSMEM 9095 9115 POTENTIAL.  
 FT TRANSMEM 9115 9135 POTENTIAL.  
 FT TRANSMEM 9135 9155 POTENTIAL.  
 FT TRANSMEM 9155 9175 POTENTIAL.  
 FT TRANSMEM 9175 9195 POTENTIAL.  
 FT TRANSMEM 9195 9215 POTENTIAL.  
 FT TRANSMEM 9215 9235 POTENTIAL.  
 FT TRANSMEM 9235 9255 POTENTIAL.  
 FT TRANSMEM 9255 9275 POTENTIAL.  
 FT TRANSMEM 9275 9295 POTENTIAL.  
 FT TRANSMEM 9295 9315 POTENTIAL.  
 FT TRANSMEM 9315 9335 POTENTIAL.  
 FT TRANSMEM 9335 9355 POTENTIAL.  
 FT TRANSMEM 9355 9375 POTENTIAL.  
 FT TRANSMEM 9375 9395 POTENTIAL.  
 FT TRANSMEM 9395 9415 POTENTIAL.  
 FT TRANSMEM 9415 9435 POTENTIAL.  
 FT TRANSMEM 9435 9455 POTENTIAL.  
 FT TRANSMEM 9455 9475 POTENTIAL.  
 FT TRANSMEM 9475 9495 POTENTIAL.  
 FT TRANSMEM 9495 9515 POTENTIAL.  
 FT TRANSMEM 9515 9535 POTENTIAL.  
 FT TRANSMEM 9535 9555 POTENTIAL.  
 FT TRANSMEM 9555 9575 POTENTIAL.  
 FT TRANSMEM 9575 9595 POTENTIAL.  
 FT TRANSMEM 9595 9615 POTENTIAL.  
 FT TRANSMEM 9615 9635 POTENTIAL.  
 FT TRANSMEM 9635 9655 POTENTIAL.  
 FT TRANSMEM 9655 9675 POTENTIAL.  
 FT TRANSMEM 9675 9695 POTENTIAL.  
 FT TRANSMEM 9695 9715 POTENTIAL.  
 FT TRANSMEM 9715 9735 POT



RA MEDLINE=97426617; PubMed=9278503;  
RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474(1997).  
RN [3]  
RP CHARACTERIZATION, AND GENE NAME.  
RX MEDLINE=95362656; PubMed=7635817;  
RA Thoeny-Meyer L., Fischer F., Kunzler P., Ritz D., Hennecke H.;  
RT *Escherichia coli* genes required for cytochrome c maturation.";  
RL J. Bacteriol. 177:4321-4326(1995).  
CC -1- FUNCTION: REQUIRED FOR THE BIOSYNTHESIS OF C-TYPE CYTOCHROMES.  
CC POSSIBLE SUBUNIT OF A HEME LYASE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCD1/NRFE/CCSA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U00008: AAA16388.1: -  
DR EMBL: AE000309: AAC75256.1: -  
DR Ecogene: E612054: ccmf.  
DR InterPro: IPR002541: Cyto\_c\_asm.  
DR InterPro: IPR003567: Cyto\_c\_biom.  
DR Pfam: PF01578: Cyto\_c\_asm.1.  
DR Cytochrome c-type biogenesis; Transmembrane; Inner membrane;  
KW Complete proteome.  
FT TRANSMEM 9 29 POTENTIAL.  
FT TRANSMEM 42 62 POTENTIAL.  
FT TRANSMEM 94 114 POTENTIAL.  
FT TRANSMEM 121 141 POTENTIAL.  
FT TRANSMEM 176 196 POTENTIAL.  
FT TRANSMEM 211 231 POTENTIAL.  
FT TRANSMEM 236 256 POTENTIAL.  
FT TRANSMEM 274 294 POTENTIAL.  
FT TRANSMEM 312 332 POTENTIAL.  
FT TRANSMEM 351 371 POTENTIAL.  
FT TRANSMEM 390 410 POTENTIAL.  
FT TRANSMEM 425 445 POTENTIAL.  
FT TRANSMEM 448 468 POTENTIAL.  
FT TRANSMEM 482 502 POTENTIAL.  
FT TRANSMEM 609 629 POTENTIAL.  
SQ SEQUENCE 647 AA; 71389 MW; 162316A1ED156DAB CRC64;

Query Match 4.9%; Score 118; DB 1; Length 647;  
Best Local Similarity 21.4%; Pred. No. 1.1;  
Matches 112; Conservative 75; Mismatches 158; Indels 178; Gaps 30:

OY 23 LHRIGSLRA---WRASSQLWV---SEALGFLAVYIGAPVPSSALGICLAA----- 72  
DB 22 YPLMGVANGDARMMASSRLFMFLMSVAGAFVLIV---NAFVNFTVYVVASNSNT 77  
OY 73 -----IAAYVALLSLDIDLRQATPIHMLVLYMGVDAITGSPV-----RAALV 119  
DB 78 QLPVYTRVAATG-----AHESLLMLVLLMSGWTFVAIAIFQRIPLDIVAAYLAIM 129  
OY 120 GLAKTLVLYV-----PALAAYLRNPLR---SLFSVYVITSLFVSYGIN 164  
DB 130 GWAVSGELLFLIFETSNPESRKLPNFPICRDL-NPLLDGPIILFPLP---LYMGVGS 185  
OY 165 -GMIVGVEELAVWDNRNSADTSRYVSLGNPNILAAVLVPT-----TAFSAAGW 217  
DB 166 VAFAPAIASLSGLRDSYVART-----RPWTIAAMIFLTIGIVLGSMAVYELG-W 236

OY 218 RGM-----LPKLLAIA-----AWGASLCLITY-SRG 244  
DB 237 GGMWMDVEVNASFPMLVGTALMNSLAVTQORSFKWTLILAISAFSLCLLGTPLVRS 296  
OY 245 GML-----GVAMIFVALLG-LYWRQPR-----LPAPWR---WLPV 279  
DB 297 GVLVSVAHFASDPARGMFLAFMVLVIGSLLEFARGHKVRSRVNMLWSRESILLANN 356  
OY 280 VLGGVAVLLVAVLG-LEPLAVRVL---SIVFGREDDSSNNRINWMLAVLOMDRPLG 335  
DB 357 VL-LVAALVLYLGLTLLPLVHKQGLGSLISIGEPFNTMF---TWL---MVFALLG 407  
OY 336 IGP-----GNTAFNLVYPLQOARFALSAYSVPLEAVEGGL 374  
DB 408 VQPLVRMGDRPKRKLNLLIATFISTLVLSLLPMLPESKVAAT-----VL 454  
OY 375 GELTAFMALLVYAVVAVROVSRLRRDQAFW-LMASIAGIA 416  
DB 455 GLAMACWIAVLAIAEAALRIS--RGTKTFTSYGWVAHGLIA 495

RESULT 12  
Y4NK\_RHISN STANDARD; PRT; 662 AA.  
ID Y4NK\_RHISN  
AC P55583;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DE HYPOTHEITICAL 72.8 KDA PROTEIN Y4NK.  
GN Y4NK.  
OS Rhizobium sp. (strain NGR234).  
OC Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizodium.  
OX NCBI\_TaxID=394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97305956; PubMed=9163424;  
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
RA Perret X.;  
RT "Molecular basis of symbiosis between *Rhizobium* and legumes.";  
RL Nature 387:394-401(1997).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: NONE OBVIOUS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AE000087: AAB91790.1: -  
RW Hypothetical protein; Plasmid; Transmembrane.  
FT TRANSMEM 10 30 POTENTIAL.  
FT TRANSMEM 46 66 POTENTIAL.  
FT TRANSMEM 68 88 POTENTIAL.  
FT TRANSMEM 101 121 POTENTIAL.  
FT TRANSMEM 167 187 POTENTIAL.  
FT TRANSMEM 193 213 POTENTIAL.  
FT TRANSMEM 217 237 POTENTIAL.  
FT TRANSMEM 263 283 POTENTIAL.  
FT TRANSMEM 285 305 POTENTIAL.  
FT TRANSMEM 312 332 POTENTIAL.  
FT TRANSMEM 342 362 POTENTIAL.  
FT TRANSMEM 373 393 POTENTIAL.  
FT TRANSMEM 394 414 POTENTIAL.  
FT TRANSMEM 432 452 POTENTIAL.  
FT TRANSMEM 460 480 POTENTIAL.  
FT TRANSMEM 485 505 POTENTIAL.  
SQ SEQUENCE 662 AA; 72842 MW; 8646516DD1A1825D CRC64;





CC STRAIN-09906;  
 RA Yang C.H., Azad H.R., Cooksey D.A.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REQUIRED FOR THE BIOGENESIS OF C-TYPE CYTOCHROMS.  
 CC POSSIBLE SUBUNIT OF A HEME LYASE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NRFE/CCSA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U04827; AAC04226.1; -;  
 DR InterPro: IPR002541; CytC\_asm.  
 DR InterPro: IPR003567; CytC\_blog.  
 DR Pfam: PF01578; CytC\_asm. 1.  
 KW Cytochrome c-type biogenesis; Transmembrane; Inner membrane.  
 FT TRANSMEM 15 35 POTENTIAL.  
 FT TRANSMEM 49 69 POTENTIAL.  
 FT TRANSMEM 99 119 POTENTIAL.  
 FT TRANSMEM 126 146 POTENTIAL.  
 FT TRANSMEM 181 201 POTENTIAL.  
 FT TRANSMEM 216 236 POTENTIAL.  
 FT TRANSMEM 250 270 POTENTIAL.  
 FT TRANSMEM 279 299 POTENTIAL.  
 FT TRANSMEM 318 338 POTENTIAL.  
 FT TRANSMEM 360 380 POTENTIAL.  
 FT TRANSMEM 395 415 POTENTIAL.  
 FT TRANSMEM 426 446 POTENTIAL.  
 FT TRANSMEM 450 470 POTENTIAL.  
 FT TRANSMEM 493 513 POTENTIAL.  
 FT TRANSMEM 618 638 POTENTIAL.  
 SQ . SEQUENCE 660 AA; 72172 MW; 3C6F4B578C56436B CRC64;

Query Match 4.8%; Score 116.5; DB 1; Length 660;  
 Best Local Similarity 20.3%; Pred. No. 1.4;  
 Matches 107; Conservative 58; Mismatches 180; Indels 181; Gaps 21;

CC 3 VMDLTFAYHPOQMGHSSFLHRLFGSL-----RAW 33  
 DB 39 LKMSLA-----QPAKMGQFAFLFAFCGLYAFKTDPSVAAYVANNNSALPWYLFSAWA 94  
 CC 34 RA-SSQLVWSEALG-----GFLAVVYGSAF-- 60  
 DB 95 GAHEGSLLLMALITLGGWTFEAVSFNSQLPQVMLARVLAVMGMSIGFLFLIMTSNPFER 154  
 CC 61 ----VSSALGGLALIAIYMALISTDIDLKQATPIHMLVLLYGVCLALATGLSPVRAA 116  
 DB 155 MLPQIADHDHNP-----LDIGLIVHPP-----MLYMG----- 185  
 CC 117 ALVGLAKLTLVLLVFLAARVLKRPRLR-----SLLEVVYITSFVSYGILNQIYGV 170  
 DB 186 -YVG-----FSYAPFAIALALGGLDAAMARMSRPMTIVAMAFIGISITIGSSWAAIYE 238  
 CC 171 EELATVVDNRNSVAADFTSRVSYLGNPMLAAVLVPTTAFSAAI-----GVNKGMLPKLIA 226  
 DB 239 LGMGWMPFMDPVENAS-----FMPMLVGTALHLSLAIVPEKRGVPSW-TVLLA 285  
 CC 227 IATGASSICLLITTSRGGMISFVAMIFVALLGLYFQPRLPAPRRRLFPVYLGLIVA 286  
 DB 286 IAFSLTLLAALVRS-----GVLSVNAFA-----SDPER-----GVFTL 321  
 CC 287 VLLVAVIGLEPLRVYLSTFVGRDSSNNFRINWMLAVLQMIODRPMGLIGPNTAFNLV 346  
 DB 322 IFLLFVVG-----GSLTIFVAPVYKSOVGFNIMSRRETLILGNNLVLVAAASMTLLGLT 375  
 CC 347 YPLVQOARFTALSAYSVPLEVAVEGGLGLTAFAMLLLVTAATAVAVOVSRLRRDRNQAF 406

DB 376 YPLVLDALSAKLSVGPPEFNLFIPLMGJ-----LMVYNAAGV-----LYRWKDTPVK 424  
 CC 407 WLM-----ASLAGLGMIGHGLEDPTVLYRPEASTLMLICIG 442  
 DB 425 WLVGMALAPVLLGSVALAVAG--IAYGDFNMVAVLATPLLAAMVILIAG 469

RESULT 15  
 LAT2\_RAT  
 ID LAT2\_RAT STANDARD; PRT; 533 AA.  
 AC Q9WVR6;  
 DT 20-AUG-2001 (Rel. 40; Created)  
 DT 20-AUG-2001 (Rel. 40; Last sequence update)  
 DT 20-AUG-2001 (Rel. 40; Last annotation update)  
 DE LARGE NEUTRAL AMINO ACIDS TRANSPORTER SMALL SUBUNIT 2 (L-TYPE AMINO  
 DE ACID TRANSPORTER 2).  
 GN SIC7A8 OR LAT2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Small intestine;  
 RX MEDLINE=99321902; PubMed=10391916;  
 RA Segawa H., Fukasawa Y., Miyamoto K., Takeda E., Endou H., Kanai Y.;  
 RT Identification and functional characterization of a Na<sup>+</sup>-independent  
 RT neutral amino acid transporter with broad substrate selectivity.";  
 RL J. Biol. Chem. 274:19745-19751(1999).  
 CC -1- FUNCTION: SODIUM-INDEPENDENT, HIGH-AFFINITY TRANSPORT OF LARGE  
 CC NEUTRAL AMINO ACIDS. HAS HIGHER AFFINITY FOR L-PHENYLALANINE THAT  
 CC LAT1. L-ALANINE IS TRANSPORTED AT PHYSIOLOGICAL CONCENTRATIONS.  
 CC PLAYS A ROLE IN BASOLATERAL (RE)ABSORPTION OF NEUTRAL AMINO ACIDS.  
 CC SUBUNIT: DISULFIDE-LINKED HETERODIMER WITH THE AMINO ACID  
 CC TRANSPORT PROTEIN SIC3A2/4F2HC.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MANY AMINO ACID  
 CC PERMEASES. GLYCOPROTEIN-ASSOCIATED AMINO ACID TRANSPORTERS (GPAT)  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB024400; BA082517.1; -;  
 DR InterPro: IPR002293; AA\_rel-permease\_1.  
 DR InterPro: IPR002027; Amino\_acid\_permease.  
 DR Pfam: PF00324; aa-permeases; 1.  
 KW Transport; Amino-acid transport; Transmembrane.  
 FT TRANSMEM 41 61 POTENTIAL.  
 FT TRANSMEM 73 93 POTENTIAL.  
 FT TRANSMEM 114 134 POTENTIAL.  
 FT TRANSMEM 156 176 POTENTIAL.  
 FT TRANSMEM 190 210 POTENTIAL.  
 FT TRANSMEM 232 252 POTENTIAL.  
 FT TRANSMEM 269 289 POTENTIAL.  
 FT TRANSMEM 311 331 POTENTIAL.  
 FT TRANSMEM 363 383 POTENTIAL.  
 FT TRANSMEM 389 409 POTENTIAL.  
 FT TRANSMEM 423 443 POTENTIAL.  
 FT TRANSMEM 448 468 POTENTIAL.  
 SQ . SEQUENCE 533 AA; 58190 MW; 99479DB60DA69DF0 CRC64;

Query Match 4.8%; Score 116; DB 1; Length 533;  
 Best Local Similarity 22.5%; Pred. No. 1.3;  
 Matches 85; Conservative 54; Mismatches 116; Indels 122; Gaps 18;

[illegible]

Search completed: January 10, 2002, 23:09:02  
Job time: 304 sec

OM protein - protein search, using sw model

Run on: January 10, 2002, 23:04:08 ; Search time 56.67 Seconds

1205.385 million cell updates/sec

Title:	US-09-887-038-3
Perfect score:	2426
Sequence:	1 MIVWQTLTFAHYQPOOWGHS.....WQPPSKQLPPEAEHSEDEKM 467

Scoring table: BLOSUM62

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum match 0%

Listing first 45 summaries

Database : SPTREMBL\_17:\*

```

1:  SP archaea: *
2:  sp_bacteria: *
3:  sp_fungi: *
4:  sp_human: *
5:  sp_invertebrate: *
6:  sp_mammal: *
7:  sp_mbc: *
8:  sp_organalle: *
9:  sp_page: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2426	100.0	457	2	P72547	P72547 synechococcus
2	1132	46.7	474	2	P73958	P73958 synechocystis
3	206	8.5	438	2	O9kth1	O9kth1 vibrio chol
4	201.5	8.3	387	2	P72672	P72672 synechocystis
5	195	8.0	443	2	O9rth43	O9rth43 xylella fast
6	180	7.4	418	2	O9rth9	O9rth9 pseudomonas
7	178	7.3	611	2	O9ABR1	O9ABR1 caulobacter
8	172.5	7.1	438	2	O9KRG4	O9KRG4 streptomyces
9	170	7.0	391	2	O9RKE4	O9RKE4 pseudomonas
10	169.5	7.0	384	2	O9RAZ8	O9RAZ8 deinococcus
11	168	6.9	422	2	O9A5L8	O9A5L8 caulobacter
12	166.5	6.9	557	2	O9KRW6	O9KRW6 vibrio chol
13	161.5	6.7	881	2	O9A4U5	O9A4U5 caulobacter
14	159	6.6	440	2	O06944	O06944 synechocystis
15	154	6.3	360	2	O9XC97	O9XC97 klebsiella
16	151.5	6.2	401	2	O9HUG6	O9HUG6 pseudomonas
17	151.5	6.2	461	2	O9rTE8	O9rTE8 pseudomonas
18	149	6.1	395	2	O9Z404	O9Z404 pseudomonas
19	149	6.1	539	2	P74814	P74814 sphingomonas

20	144.5	6.0	381	2	Q9G309	Q9G909	thiobacillus
21	139.5	5.8	368	2	Q9R725	Q9R725	delnoccocus
22	139.5	5.8	407	2	Q9A5Z6	Q9A5Z6	caulobacter
23	138.5	5.7	402	2	Q9Z1T8	Q9Z1T8	eschericchia
24	138	5.7	515	2	Q9RRT1	Q9RRT1	delnoccocus
25	137.5	5.7	443	2	Q9KEH4	Q9KEH4	bacillus
26	137.5	5.7	474	2	Q9R0Y9	Q9R0Y9	delnoccocus
27	134	5.5	335	2	Q9RSL2	Q9RSL2	vibrio
28	134	5.5	402	2	Q9I2B6	Q9I2B6	pseudomonas
29	133.5	5.5	535	2	Q9R9W6	Q9R9W6	pseudomonas
30	133.5	5.5	585	2	P726Z7	P726Z7	synecocyst
31	131.5	5.4	515	2	Q9RWT0	Q9RWT0	strepomyces
32	131	5.4	524	10	Q809O5	Q809O5	arabidopsis
33	130.5	5.4	577	2	Q9L134	Q9L134	streptomyces
34	130	5.4	536	2	Q9I6A1	Q9I6A1	pasteurella
35	130	5.4	669	2	Q9KE20	Q9KE20	bacillus
36	129.5	5.3	405	2	Q68200	Q68200	aeropyrum
37	129	5.3	487	1	Q9YGV7	Q9YGV7	halobacter
38	128.5	5.3	530	2	Q9I7Z6	Q9I7Z6	delnoccocus
39	128.5	5.3	566	1	Q9HS12	Q9HS12	halobacter
40	128	5.3	451	2	Q515Z4	Q515Z4	pseudomonas
41	127.5	5.3	437	2	Q9I0U5	Q9I0U5	pseudomonas
42	127.5	5.3	678	2	Q9A4E2	Q9A4E2	caulobacter
43	127	5.2	591	2	Q531S0	Q531S0	mycobacter
44	127	5.2	709	2	Q9RYK7	Q9RYK7	delnoccocus
45	126.5	5.2	467	2	Q311Y9	Q311Y9	pseudomonas

## ALIGNMENTS

RESULT	1			
PT2547				
ID	P72547	PRELIMINARY:	PTT:	467 AA.
AC	P72547.			
DT	01-FEB-1997 (TREMBLrel. 02.			Created)
DT	01-FEB-1997 (TREMBLrel. 02.			Last sequence update)
DT	01-AUG-1998 (TREMBLrel. 07.			Last annotation update)
DE	PUTATIVE MEMBRANE PROTEIN.			
GN	DC14.			
OS	Synechococcus sp. (strain PCC 7942) (Anacyctis nivalans R2)			
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.			
OX	NCBI_TaxID:1140;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-PCC 7942;			
RA	Boofil D.J., Lisman-Hurwitz J., Ronen-Tarazi M., Kaplan A.;			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
DR	EMBL, U62616; AAB08477.1;			
SO	SEQUENCE	467 AA:	51570 MW:	2C3FEB404ED462BF CRC64;

Query Match	100.0%;	Score 2426;	DB 2;	Length 467;
Post Local Similarity	100.0%;	Score 2426;	DB 2;	Length 467;

Best local similarity 100.0%; Pired. NO. 2.2e-123;  
Matches 467; Conservative 0; Mismatches 0; Indels 0;  
Gaps 0;

QY	1	MTWQTLTFAHYAPQPOOMGHSFLEHRLFGSLRAMRASSQLLWVSEALGGFLLAIVVGSAPF	60
Db	1	MTWQTLTFAHYAPQPOOMGHSFLEHRLFGSLRAMRASSQLLWVSEALGGFLLAIVVGSAPF	60
QY	61	VPSSALGIGLALAAAYVMALSLTDIDLRQKPTPHMLVLLYWGVDALATGSLSPVRAALVG	120
Db	61	VPSSALGIGLALAAAYVMALSLTDIDLRQKPTPHMLVLLYWGVDALATGSLSPVRAALVG	120
QY	121	LAKTLTYLLVFLAARVLRNRLSLFESVVYITSLFVSVYGINOMYIGVEFLATWMDRN	180
Db	121	LAKTLTYLLVFLAARVLRNRLSLFESVVYITSLFVSVYGINOMYIGVEFLATWMDRN	180
QY	121	LAKTLTYLLVFLAARVLRNRLSLFESVVYITSLFVSVYGINOMYIGVEFLATWMDRN	180
Db	121	LAKTLTYLLVFLAARVLRNRLSLFESVVYITSLFVSVYGINOMYIGVEFLATWMDRN	180
QY	181	SVADPFSRYSYSLGNPNLLAAVLYPTTAFSAALIGVWRGMKLLAIAATGASSCLILT	240
Db	181	SVADPFSRYSYSLGNPNLLAAVLYPTTAFSAALIGVWRGMKLLAIAATGASSCLILT	240
QY	241	YSRGGVLGFVAMIFVWALLGYMFPQRLPAPAWRRMLFVVLGGIAVAVLLVAVLGLPIRLV	3000
Db	241	YSRGGVLGFVAMIFVWALLGYMFPQRLPAPAWRRMLFVVLGGIAVAVLLVAVLGLPIRLV	3000

```

|||||
Db 241 XSRGMLGFVAMIFWALLIGYFQPRLPAPRRHLFPVYLGIVALLVAVLGEPLKRV 300
QY 301 RVLSTFEVREDESSNNFRINWMLAVLQMDRPMIGIGGNAPFNIVYLYQOARFALISA 360
Db 301 RVLSTFEVREDESSNNFRINWMLAVLQMDRPMIGIGGNAPFNIVYLYQOARFALISA 360
QY 361 XSVPLEVAVEGGLGLTFAPFAMLLVTAVTAAROVSRLRRDNPQAFMLASLAGMLG 420
Db 361 XSVPLEVAVEGGLGLTFAPFAMLLVTAVTAAROVSRLRRDNPQAFMLASLAGMLG 420
QY 421 HGLFDFTVLYRPEASTLWMLCIGALASFQWQOPQSKOLPPEASHSDEKM 467
Db 421 HGLFDFTVLYRPEASTLWMLCIGALASFQWQOPQSKOLPPEASHSDEKM 467

RESULT 2
P73958 PRELIMINARY; PRT; 474 AA.
AC P73958;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOTHEICAL 52.2 KDA PROTEIN.
GN SLR1515.
OS Synecchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
DR EMBL: D90911; BAA18025.1; -.
DR InterPro: IPR001457; Oxidored_q3.
DR Pfam: PF00499; Oxidored_q3.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 474 AA; 52241 MW; 017A49BC2758EC6 CRC64;

Query Match 46.7%; Score 1132; DB 2; Length 474;
Best Local Similarity 47.5%; Pred. No. 8.2e-54;
Matches 223; Conservative 79; Mismatches 163; Indels 4; Gaps 2;

QY 1 MTWOTLTFAHQPOQMGHSEFLHRLFGSLRAMRASQQLVWSEALGGFLAVVYGSAPF 60
Db 5 ISIMWSLMEGGSGPQEMRGSVLHRLVGMQSGMIOASVLMHPEFLGATLVAIIIFIAFP 64
QY 61 VPSSALGIGLAAIAAYWALLSLTDIDLRQATPIHMLVLLYMGVDALATGLSPVRAALVG 120
Db 65 TSTMLGIFMLCGAFMLLTFADPGKGLPIHVLVFAFYCISAIAGFGSVKMAASG 124
QY 121 LAKTLVLYLVLFAARVLRNPRLRSLSVYVITSLFVSYGLNOMIYGVIELATWVRN 180
Db 125 LAKTLVLYLVLFAARVLRNPRLRSLSVYVITSLFVSYGLNOMIYGVIELATWVRN 184
QY 181 SAVDETSRYSYLGNPMLLAAYLVPTAFSAALGVNRGMLPKLAIATGASSICLITLY 240
Db 185 SFLAQTATRYSPFLGNPLLAAYLVPTMGSLASLVVWRMRMKLLGATMVIVNLCLFET 244
QY 241 YSRGMLGFVAMIFWALLIGYFQPRLPAPRRHLFPVYLGIVALLVAVLGEPL 298
Db 245 OSRGMLAVLAVGATFLLCYFWMLPOLPKFQWRMSLPLATA--VAVILGGALLAVBEI 302
QY 299 RRVVLSIFVGRDESSNNFRINWMLAVLQMDRPMIGIGGNAPFNIVYLYQOARFAL 358

```

```

Db 303 RLRAMSIFAGREDESSNNFRINWMLAVLQMDRPMIGIGGNAPFNIVYLYQOARFAL 362
QY 359 XSVPLEVAVEGGLGLTFAPFAMLLVTAVTAAROVSRLRRDNPQAFMLASLAGMLG 418
Db 363 XSVPLEVAVEGGLGLTFAPFAMLLVTAVTAAROVSRLRRDNPQAFMLASLAGMLG 422
QY 419 LGHGLFDFTVLYRPEASTLWMLCIGALASFQWQOPQSKOLPPEASHSDEKM 467
Db 423 LVHGMVDTWVYRPPVSTLWMLVLAIVASQWASAOARLEASKENEDKPL 471

RESULT 3
Q9KTH1 PRELIMINARY; PRT; 438 AA.
ID Q9KTH1;
AC Q9KTH1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHEICAL PROTEIN VC0931.
GN VC0931.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=EL TOR N16961 / SEROTYPE O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gilm M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ullrich J., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL: AE004176; AAF94093.1; -.
DR TIGR: VC0931; -.
KW Complete proteome.
SQ SEQUENCE 438 AA; 49177 MW; FDF34AAC184C429B CRC64;

Query Match 8.5%; Score 206; DB 2; Length 438;
Best Local Similarity 24.7%; Pred. No. 0.00046;
Matches 120; Conservative 68; Mismatches 200; Indels 98; Gaps 19;

QY 5 OTLFAHQPOQMGHSEFLHRLFGSLRAMRASQQLVWSEALGGFLAVVYGSAPF 60
Db 6 KTLFPA-----FLFVFWLPPIPLGSNRPW-----AWSLNEIYVALL----- 42
QY 57 SAPVPSALG-LGLAAIAAYWALLSLTDIDLRQATPIHMLVLLYMGVDALATGLSPVRA 115
Db 43 -----LGCLEFLPBAQCMLAIQRAKYL--IYVALFTLMSWMQSPISGLSDQT 90
QY 116 AALVGLAKLILYLVFLAARVLRNPRLRSLSVYVITSLFVSYGLNOMIYGVIELATWVRN 162
Db 91 LVFISAKSLSHYLOICVLAALVDVTPHKLMLATMTMASGICAFYAGVITLLELQTSPE 150
QY 163 -----LNMIVGVEELATWVRNSVADTSRYSYLGNPMLLAAYLVPTAFSAALGVNR 218
Db 151 FSLPELRASG-----SFVYHNHIANFL-KLNCLGIGLLIAELINQTT-----GWR 197
QY 219 GMLPKLAIATGAS-----SLCLITYSRGMLGFVAMIFWALLIGYFQPRLP 269
Db 198 NTLKLLAILLSDKAFIRGLGIVIVLAVLTVRSRMGNIATFVALISGSLILILFYKSK-- 255
QY 270 APARMRKLPVYLGIVALLVAVLGEPLRVLSTFEVREDESSNNFRINWMLAVLQMDR 329
Db 256 -PKSLYLLIYSLFTVLDLVASVNMFGDKVRQR-----LAFTSLQNESRDVVDALNAIQ 309
QY 330 DRPMIGIGGNAPFNIVYLYQOARFAL--SAVSVPLEVAVEGGLGLTFAPFAMLLVTVA 387

```

Db 310 ERLPLGF--GNGCTFSTPESSYNOGEVALFYDHAANDVQFTLESLGATSVFGIMVLFCA 367  
QY 368 VTAAQVSRRLRRDRPQAFWLMASLAGMIGHLFTVLYRPAASLTMWLCITAINSF 447  
Db 368 WRAL-SAFRRNRINKMKRGIGL-GSLMAITIGMLHMSVDFPQAP-ATYALEFCELLIANW 424  
QY 448 WQOPPS 453  
Db 425 SMTLPT 430

RESULT 4  
P72672 PRELIMINARY: PRT: 387 AA.  
AC P72672:  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE HYPOTHEtical 43.0 KDA PROTEIN.  
GN SLR0728.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_Taxid=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97061201; PubMed=8905231;  
RA Kaneke T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res 3:109-136(1996).  
DR EMBL; D90899; BAI1674.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 387 AA; 42984 MW; 81D6C8DFCA25A1C9 CRC64;

Query Match 8.3%; Score 201.5; DB 2; Length 387;  
Best Local Similarity 25.98; Pred. No. 0.00071;  
Matches 103; Conservative 54; Mismatches 179; Indels 61; Gaps 17;  
QY 95 WLVLLYMGVDALATGSLSVRAAALVGLAKLTLYLVFALAAVRNPRLSILFSVYVIT 154  
Db 2 WAALLTWPL--LTTHADRGEALLGLANFPYPTVFLAFSQVLCQFQQLTLMMLTAN 59  
QY 155 SLFVSVYGLNMIYGEELATW--VDRNSVADF--TSRVSYLGNPNILAAVYPTTAF 209  
Db 60 TVVLIIIGFGQ-VYGGWATPMMLVAIGTLVAGRPCEGRMSLLMYANLFSAMILLMPPL 118  
QY 210 S-AAAGVGRW-----LPRLL--ATAAGASSLCILLYSRGGMGFVAMITW 256  
Db 119 SLGLTIOSVRWHHTAPSNNRFLNPLVALLVACILAEIALVLTGSRSAW-GIGILLGIA 177  
QY 257 ALLGLYFQPLPAPWRMLFPVYGLVAVLLVA--VLGLEPLRAYVLSPVGR--- 309  
Db 178 YAYILSWY-----WLV-ALAGATFANVLASGEPPEKEPLRQVYPTFPGRLSDE 226  
QY 310 ---EDSSNNFRINWLAVALOMIDRPMWIGIPGNTAFNLYPLVQAQAFALIS-AYSVL 365  
Db 227 LYPRRYRFLALSTQWOFQMDMFLDQPIFGQGLRNPT-----PLYOAAANVWIGPHNLVL 281  
QY 366 EVAVEGGILG---LTAAMLLVTAVAVROVSLRSD---RNPQAFWLMASLAGIAG 417  
Db 282 MMLGETGLTALMAGAAVGLAOGMVLIT---HLSSGGGFRRRSOHLLLSYGIAFAA 337  
QY 418 MIGHGLFDTVLYRPAASLTMWLCITAINSFQOPPS 454  
Db 338 LCIVNLFDVYITFDMRNWVIGWIFLAALAGVSGQRGTR 374

RESULT 5  
Q9PH43 PRELIMINARY: PRT: 443 AA.  
ID Q9PH43:  
AC Q9PH43:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MEMBRANE PROTEIN.  
GN XE0103.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_Taxid=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
RA Alvares R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,  
RA Barros M.R., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,  
RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
RA Gantier M., Goldman G.H., Goldman H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Homelisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.P.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marques C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
RA Nianai A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pezquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., de Silva A.M., de Silva A.R., Silva W.A. Jr.,  
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsunako M.H.,  
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
RL Nature 406:151-159(2000).  
DR EMBL; AEO03864; AAF82916.1; -  
KW Complete proteome.  
SQ SEQUENCE 443 AA; 48239 MW; E53A1CE9A66FB56F CRC64;

Query Match 8.0%; Score 195; DB 2; Length 443;  
Best Local Similarity 25.28; Pred. No. 0.0018;  
Matches 107; Conservative 47; Mismatches 158; Indels 112; Gaps 21;  
QY 56 GSAPFVSSALGIGIAIAVW-----ALSTPDIIDRQATPIHMLLYMGVDALA 107  
Db 63 GGTGLDLSPMALNLSVLFAMVWLPQVSLPGALLESLSLKAVNTDKLYLFFMW-LCAIA 120  
QY 108 TGLSPVRAALVGLAKLTLYLVFALAAVRNPRLSILFSVYVITSLFVSYGLNMI 167  
Db 121 VGSPEERERTEFGGLAVIGLITWTDALQAAMLHT-----SPLTWSINQKQAI 167  
QY 168 YGVEELATWDRNSVAQFTRSYVSYLGNPNILAAVYPTT-----AFSAAGVWNGWL 221  
Db 168 SGYFGCT--RQMT--LADRLSGTFPCWIKGQTIATLSPPVLAQMORGLALW---- 218  
QY 222 PKLDAIATGASSLCILLYSRGMLGF--VAMIFWALLGTFQPLPAPRRWLPV 279  
Db 219 --LLVVVVLG---IVLLASRASWITYALLIMLSGRVIGI-----RRL--- 259  
QY 280 VLGGLVAVLVAVVIGLEPL--RV-RVLSIFVGR--EDSSNNFRINWLAVALOMIDR 332  
Db 260 GMGALFLPLALAVIAFSPPQTRRIDRTAAVAFADHAGAGVDALSGRSQIWMQALCMIOAHP 319

```

OY 333 WLIGEGMTAFNLVLP-----LYQARFALSAYSVPLEVAVGGILGL-----TA 378
DB 320 LSGGVV--RGFRDAYPCNPTEPERIPAMGAGALHAIQVLEILSBTVIGILLMIAGAA 377
OY 379 FAWLLLVTAVTAVROYSRRLRRDRNPA-----FW--LMAISLACL- 415
DB 378 MM-----RAMRYATAARDRARPAMISLTLATVPFNTHLAFSSFWGALLMLAGLY 430
OY 416 AGML 419
DB 431 AGAL 434

RESULT 6
O91M9 PRELIMINARY: PRT: 478 AA.
AC O91M9:
DT 01-MAR-2001 (TREMBlrel, 16, Created)
DT 01-MAR-2001 (TREMBlrel, 16, Last sequence update)
DE 01-JUN-2001 (TREMBlrel, 17, Last annotation update)
DE HYPOTHETICAL PROTEIN PA2240.
GN PA2240.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laibig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL: AE004649; AAG05628.1;
DR InterPro: IPR003439; ABC_TRANSPORT; UNKNOWN.1.
DR PROSITE: PS00211; ABC_TRANSPORTER; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 478 AA: 52495 MW: 604F36DABAF70F1 CRC64;

```

Query Match 7.4%; Score 180; DB 2; Length 478;  
 Best Local Similarity 23.0%; Pred. No. 0.013;  
 Matches 124; Conservative 67; Mismatches 159; Indels 188; Gaps 29;

```

OY 48 GFLAVVYVYSA-----PFVPSALGLIAATA-----YMAIL-----SL 82
DB 4 GFLVAVVYVGLAFGALLMLMSPAKAGAA--VGLAAAVTILRRPFWCLLFLAVATPMPYST 62
OY 83 TDIDLRQATPIHWLVLYWGV---DALATGSPVBAALVGLAKTLYLVFA----- 132
DB 63 VNWGIR-STVSEAILATLWGAIVLWHIFLSRLPPAPALRYRSTERMLLMFTVLPEVYG 121
OY 133 -----LAARVL-----RNPRLSLISFVVVITSLFVSYY- 161
DB 122 QVSTAKASGLSNLRLNLSIYFVLAAGLVERKNRRETLVIALTLGLTAMLMISIAVEI 181
OY 162 -----GLNOMIYVEELATWVDNRNSVADFTSRVSYSLGNPNL---LA 200
DB 182 RYRSASGAPILAMFNVANIDTLKRGLEAL-----SSMGSPMWHHPNATGIM 229
OY 201 AYLVTTFPSAAAGVWFGMLPKLLAATGASSICLLITYSRGGWLGFMATIFWALLG 260
DB 230 ALLPLACVGYANGWRRGIG--LAVAVLGAAA--ILLASSRGAMSLAVVLF--WMSLR 284
OY 261 LYWFQPRLLPAPRRMLFPVYLG--GLVAVLVAVGLPELRVYLSIFVGRDSSNNPRIN 319

```

```

DB 285 KVPYGRLL-----LLLGVLVVL---VLSPIQLERLATIF--SPONASTEVRFD 330
OY 320 VWLAVLQMIODRPMWLIG-----PG-----NTFANLYPLYQARFALSAYSVPL 365
DB 331 EYRMFPKAVARYP--LGISFKVDPPVPGTDLIGISNLMINFWY----- 371
OY 366 EVAVEGGILGLTAPFAMLLVTAVTAVROYSRRLRRDRNP-----QAFWLMASLAGML 419
DB 372 KVGL--GGMFLFLATVW-----RWMREARPEKGPILRTRONALIML--GSTGIIAAL 419
OY 420 GHGLFDIVL--VRPEASTIMWLICG-----AIASFWQPOP-----SKOLPEPE 461
DB 420 VSGLEDFHYSFAVVMIGLFWLLVGINLLEARRLPERPQOPRAVGVRKLRKERGAE 477

RESULT 7
O9ABR1 PRELIMINARY: PRT: 641 AA.
AC O9ABR1:
DT 01-JUN-2001 (TREMBlrel, 17, Created)
DT 01-JUN-2001 (TREMBlrel, 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel, 17, Last annotation update)
DE HYPOTHETICAL PROTEIN CC0165.
GN CC0165.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Meddock J.R.,
RA Potocka A., Nelson W.C., Newton A., Stephens C., Phdke N.D., Ely B.,
RA Deacy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kotonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Kuterback T., Tran K., Wolf A., Yamthuevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005690; AKR2152.1;
DR TIGR; CC0165;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 641 AA: 67175 MW: D8FF63BE76B565F9 CRC64;

```

Query Match 7.3%; Score 178; DB 2; Length 641;  
 Best Local Similarity 23.1%; Pred. No. 0.021;  
 Matches 115; Conservative 80; Mismatches 193; Indels 110; Gaps 23;

```

OY 18 GHSSFHLRLFGSLRAMRASSQLLWSE--ALGGFLAVV-----YGSAPVPSSALGIL 70
DB 10 GSSS-----FGS-----SGSRRLKVSQVAGAIALLTILIAEITAFGSELAVSSA-----F 55
OY 71 AIAIYVALLSLDID-----LRQ--ATPIHWL-----VLLYGVDAIATG-----L 110
DB 56 AALHALPPIVYLATCGWARRAARQGVSPTLMPFAIIMLTAVAMLPFGCGHPYWLXY 115
OY 111 SPVRAAALVGLAKTLYL-----VFALARYL--RNPRLSLISFVVVITSLFVSYYG 162
DB 116 EPNCGALTVDRSSLLMLVRLVGLACFLFLAQAIIIGSSERTERRALFWSLVALGAY--AIIA 174
OY 163 LNMWIYVEELATWVDNRNSVADFTSRVSYSLGNPNLAAATVPTTAFSAAGVW----- 217
DB 175 ILQHV--GVRA-----SERITATLLSNTLATILMGAVFAAMFVTOFORAG 220
OY 218 -KGMPLRL-----IAIATGASSICLLITYSRGGWLGFM--AMIFWALLGLYWPQRL 268
DB 221 GRSLIDKLPLDAGISLALLVAVFALAFGTSGRGVFATVVALAIFLIWYL----- 271
OY 269 PAPRRMLFPVYGGIYAVLVAVIGLEPLRVYLSIFVGRDSSNNPRINWVLAVIQMI 328
DB 272 -AQGRARVYVWIGG--AAVLLAVGLAMRSADITTAARLENLDGDVATROTIFAHWQAF 328

```

QY	329	QDRRWLGIGPONTNFANLVLYLQY-----ARFALSAAYSPLEAVNAGGILGTA	379
Db	329	KASPMFGYGLG-----FPAYVNMVMTSETLPALFVDVRAATHNLVQWLEBAGLVGLSM	382
QY	380	AMLLVTVAVTVAVROVSRRLDRNRDQAFWMLASLGLGMLGHGLFEDVLYRPEASTMVL	439
Db	383	LGILGV-----LIMQVCAANKPGTSGALARRAIASTALVYLHGMSPDAVQA	438
QY	440	CIGAIASFQWQPSKOLP	457
Db	439	GLGALVTS-SPARRGRLP	455
RESULT	8		
Q9KYG4			
ID	Q9KYG4	PRELIMINARY;	PRT: 438 AA.
AC	Q9KYG4		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	PUTATIVE INTEGRAL MEMBRANE PROTEIN.		
GN	SCC61A.29.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
NCBI	taxid:1902;		

RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2):  
RA Oliver K., Harris D.,  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RM [2]  
RN 'SEQUENCE FROM N.A.  
RC 'STRAIN-A3(2):  
RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.,  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RM [3]  
RN 'SEQUENCE FROM N.A.  
RC STRAIN-A3(2):  
RX MEDLINE=97000351; PubMed=8843436;  
RA Kienast H., Hopwood D.A., Denapate D., Eichner A., Cullum J.,  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RM Mol. Microbiol. 21:77-96(1996).  
RA EMBL; AL356595; CAB92274.1; -  
SQ SEQUENCE 438 AA; 45017 MW; 38173961751FEFAL CRC64;

Query Match	7.1%;	Score 172.5;	DB 2;	Length 438;
Best Local Similarity	26.6%;	Pred. No. 0.029;		
Matches 119;	Conservative 63;	Mismatches 178;	Indels 87;	Gaps 23

[illegible]

OY		327	MIDRRPWLGICPGNATFAFLVPLTQQA-----RPTALSNYSVPLE	366
	:	:	:	:
Db		294	MWRREPITGVGLGK-----FPEHRDHAIALSSGSGETDGAGYNRPOLPHNNYYL	347
OY		367	VAVEGGILIGLTAF--WLLVTATAYAROVSRRLRDNPQAFTWMASLAGLA-GMLGHGL	423
	:	:	:	:
Dd		348	VLAEGELTGILLALNGSWALL-VLGRRLRAAER----AHGAVPDCAFVAGCLVMWL	400
OY		424	FDTVL--YREASTLMWLCIGAIAISFW	448
	:	:	:	:
Dd		401	TDFAVADIGGPSIVLTAVCSG-LAAWW	426

ID	Q9KKE4	PRELIMINARY:	PRT:	391 AA.
AC	Q9KKE4;			
DT	01-OCT-2000 (Tremblrel, 15, Created)			
DT	01-OCT-2000 (Tremblrel, 15, Last sequence update)			
DT	01-OCT-2000 (Tremblrel, 15, Last annotation update)			
DE	O-ANTIGEN POLYMERASE.			
GN	Wzy.			
OS	Pseudomonas putida.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OC	Pseudomonas.			
OX	NCBI_TaxID=303;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KT2442;			
RC	MEDLINE=21128631; PubMed=11207769;			
RA	Rodríguez-Herva J.J., Reniero D., Galli E., Ramos J.L.;			
RT	"Cell envelope mutants of Pseudomonas putida: Physiological characterization and analysis of their ability to survive in soil.";			
RL	Environ. Microbiol. 1:479-488(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KT2442;			
RA	Rodríguez-Herva J.J., Reniero D., Ramos J.L.;			
RL	Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AF109736; AAF1569.2; -.			
SO	SEQUENCE 391 AA; 43109 MW; 4658DF8489DE64B0 CRC64;			

Query Match	7.0%;	Score 170;	DB 2;	Length 391;
Best Local Similarity	23.1%;	Pred. No. 0.036;		
Matches 102;	Conservative 63;	Mismatches 157;	Indels 120;	Gaps 23

```

0Y 42 WSEALGGLFLLAVYVSASFVYSS-----ALGIGLAAIAAYWALLSYTIDJROAFPIH 94
Db 7 WQAMIGIGLWELWELAAIALAASNNKYQGLVFLTMLPPLVLIAMSARVYL-VQAMRRPAL 65
0Y 95 W---LVLYWGVDAIATGIS-----PYRAALVGLAKLTLELIVFALAARVLR---N 140
Db 66 WQSVLLLLLW-----SESLIAMSAPAEPMNE-----LKRLLYLIWFLAFLPQLAQGQ 113
0Y 141 PRKSLFLF--SVYVTSLEFSVYGIQMWIYEEBELATWDBNRSVADFTSRVYS--YLGNP 186
Db 114 ARIPQLLLLGLSALLIAALVSIHK---FYGVORAPLILFRLAGIGESHPILGAYVIGSA 169
0Y 197 NLLAALYVPTTAFSAALIGWRGMLPKLLAIALANGASSCLILTFYSGMNGIFAMIFW 256
Db 170 LILMIYEPERRR-----GLQLLW---LAAALCIGAFAM--LSQSGAVLALITYVMA 217
0Y 257 ALLGLYWPBPRLPAPWRMWLPVYVGLVAVLLVAVIGLEPLRYARVSIIEFGREDSNNF 316
Db 218 PL---WFDNRISR-----VSYVALLATGLAFLAVDYIAQRSSY 255
0Y 317 RINWVLAVLQMTODRPMVLGIGPN---TAEINLYPIYQOARFTALSAYSVLEAVEGG 372
Db 256 REIFRAVYVQMIAMHPWGLGIGADYEVSAVGMHRD-HTHNMFT-----HVAEMG 305
0Y 373 LUG--LTAFAMLLTVANTAVNRQVSRLRDRNPPQAFILMMSLAGIAGLGHGFLDYLYR 430

```

Db 306 LPSGLLWVWVLTFLGELVRAR-----GTLGKV-LLGFWVSTLAAQ 347

Qy 431 PEASTL-----W---WLCIG 442

Db 348 FDAASLTGTPRAEWFISLTPVG 369

RESULT 10

Q9RZAB PRELIMINARY; PRT; 384 AA.

AC Q9RZAB;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DE HYPOTHEITICAL 41.9 KDA PROTEIN.

GN DRN0046.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI\_TaxID=1299;

RP SEQUENCE FROM N.A.

RC SMRAIN-R1;

RA MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,

RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,

RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,

RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,

RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans R1.";

RL Science 286:1571-1577(1999).

DR EMBL: AE001862; AAF12372.1; .

DR TIGR: DRN0046; .

KM Hypothetical protein; Complete proteome.

SO SEQUENCE 384 AA; 41905 MW; 744B3CB599EF128 CRC64;

Query Match

Best Local Similarity 7.0%; Score 169.5; DB 2; Length 384;

Matches 113; Conservative 56; Mismatches 158; Indels 97; Gaps 24;

Qy 50 LLAAYGSAFPVPSAALGLAIAAYWALISTDIDLROATPIHMLVLLYMGVDALANG 109

Db 21 LIAIV-----PVFP-----LYLAALACLSLRTLP-----QAAR-----GVLEFFAATQILIA 64

Qy 110 LSPVAAALVGLAKL-TLYLVFALAAVLRNRP-LRSLLEFSVVVITSLPVSYGILNMT 167

Db 65 LFTAPRLISVGLAARLILILIAAYAGVYLRDSRHLRPLMGOLIT---FVTA-----WI 116

Qy 168 YGVEELATVDRNSVADFTSHV---YSYLGPNLLAAYLVPTTAFSAAGVRCMLPKI 224

Db 117 Y-----TLST-----QGVAGQERLGHYYIYISLGLVAVVAILMIYFMGAAAMRR-WPAGI 168

Qy 225 LAIATGASSLCLITLTSRSGMIGFVAMIFWMALGLYNGQRLPAPWRW-LFPVVLG 283

Db 169 LALA-----TFRAAGSRPLAIGVSLAALAFGGR-----RWVWMLPAVLW 213

Qy 284 LVAVLVAV-LGLEPLRVVLSIFVGRDSSNNFRINWVAVLQMIODRMLIGGNTA 342

Db 214 LAAPATISLVNPFKPIRLNDQTSGR-----YVMDAVSCMETSLGSGVGYGG 265

Qy 343 FNLVY-----PLYOQ-----ARPTA--ISAYSVPLEVAVEGGLGLTAPAML 382

Db 266 PVLTFLEKDGCOLPTPLQNRKIECPQOLSRSSVWLIAHNMHLLESIGISG----- 321

Qy 383 LLYAVAVAVRQVSRDRNPQAFWMAISLAGMLGHLFTPVLY--RPEASTLMWIC 440

Db 322 LLAIWVIALMRAIOL-----GDP-----FTLAIVLGTAMVAVVAVPSHFSELMWC 372

Qy 441 IGAI 444

Db 373 VGLV 376

RESULT 11

Q9A5L8 PRELIMINARY; PRT; 422 AA.

AC Q9A5L8;

DT 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DE HYPOTHEITICAL PROTEIN CC2429.

GN CC2429.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OX NCBI\_TaxID=69394;

RP [1]

RA SEQUENCE FROM N.A.

RA MEDLINE=21173698; PubMed=11259647;

RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA Deboy R.T., Dodson R.D., Durkin A.S., Gwin M.L., Haft D.H.,

RA Kolonay J.F., Smit J., Craven M.B., Knouri H., Shelly J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL: AE005912; AKK24400.1; .

DR TIGR: CC2429; .

KM Hypothetical protein; Complete proteome.

SO SEQUENCE 422 AA; 45786 MW; 656279D26E34E1F CRC64;

Query Match

Best Local Similarity 6.9%; Score 168; DB 2; Length 422;

Matches 94; Conservative 57; Mismatches 150; Indels 76; Gaps 17;

Qy 44 BALGFLIAVYGSAPVPSAALGLAIAAYWALISTDIDLROATPIHMLVLLYMGV 103

Db 50 EAASGL-----VAPFLPAGLAF-LLINGRPMNVVRV-----VRQPLILLVAVV---V 96

Qy 104 DALATGLSP---VRAALVGLAKLTYLVLPALAAVLRNRPRLSLFSVYVITSFVS 160

Db 97 ASISWSIQPDVSIKRGAVGCT--TLAGLALASFRMAELARLFAITTCFLIVASYLSL 154

Qy 161 ----YGLNOMIYGEELATVDRNSVADFTSRVSYSLGNPLLAAYLVPT-----TAFSA 211

Db 155 ALPSTGVMTLEFPAGWGMKMKNGSLMAFGACLLG-----AAALINPRAKLFLLFAG 210

Qy 212 AALGVWGMPLKLLAIAATGASSLCLITLTSRSGMIGFVAMITVWAL-----LGLYWPQ 266

Db 211 LAIGL-----VLLSOSTALASL-----LGMTLIGFVWIVORGPALG----- 248

Qy 267 RLAPRRMRLEPVLVGLVAVLVAVLEPLRVVLSIFVGRDSSNNFRINWVAVLVO 326

Db 249 -AAATW-----AGVTGALLIAEVL--LASDMLFELIGK-DATLIGRTITMAAR 295

Qy 327 MIODRPMWIGIG-----PGNTAFNLVYPLYOARPTLASVSPLEVAVEGGLGLTAFAM 381

Db 296 QIEORPMQGYGYAAVMSDKSGMPLAVINDAKRVPQHAINSMLEQWLIGLGLIAMGL 355

Qy 382 LLYTAATVAVRQVSR 398

Db 356 FYLOTMSLAVAAYVRER 372

RESULT 12

Q9K0W6 PRELIMINARY; PRT; 597 AA.

AC Q9K0W6;

DT 01-OCT-2000 (TReMBLrel. 15, Created)



DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)  
 DE HYPOTHETICAL PROTEIN VC0393.  
 GN VC0393.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.D., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL: AE004127; AAF93566.1; -  
 DR TIGR: VC0393; -  
 DR InterPro: IPR000215; Serpin.  
 DR PROSITE: PS00284; SERPIN; UNKNOWN\_1.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 597 AA; 67451 MW; D33513A00836688C CRC64;

Query Match 6.9%; Score 166.5; DB 2; Length 597;  
 Best Local Similarity 24.3%; Pred. No. 0.083;

Matches 109; Conservative 60; Mismatches 185; Indels 95; Gaps 22;

QY 49 FLAAVYSSAP--FVPS-SALUGLAATAAYALLSLT-----DIDLRQATPIHMLVLL 99  
 DB 29 FLAAVYLLAHMFEMPFGAGLASENTTWTATSTITAIQLYQLANQAIRYSKLITG 88  
 QY 100 YMGVDALATGLSPV-----RAAATVGL-AKLTLYLTFALAAVLRNRLSLFS 149  
 DB 89 LMLSCALLT--LPITYSNADWFHTGLRVLGVLAGFTILFVLQOF--RESNKKIKORLMP- 143  
 QY 150 VVYITSLFVSVGLNOMIYGEELATWDRNSVAD--TSRYSVSLGNPNLLAAVLYPT 206  
 DB 144 -ILTSVIEAVIGLIQYF-----WLEPGNPFQYDFRANRPGYFQGPQVMAWSLATG 194  
 QY 207 TAFSAAGVWRGNL-----PKLLAIAATG-ASSICLLITTSRSGWLGFPVMT 253  
 DB 195 LVLS-----GYLLARQPEKYSRMSKIGILYATPPLTAPLLVVLASRTGMLASLISI 246  
 QY 254 FWMALLGLYWFQRLPAPRRMLFPVVLGIVAVLLVAVLLEPLRVRLSTFVGRDSS 313  
 DB 247 ---ALVAPLYRFAARRRRTLMLAITLG--ICAGFTAMYG-----QGTSGFVANKADL 295  
 QY 314 NNFRLNVLAVIOMIQDRPWLGIGPGNTAFNLVPLYOQARETALSAVSVPLEV----- 367  
 DB 296 ESPRRFTPEPOTIDMTIEKFTGYGK--FEAQVLLYTAHQQLNSSYPAGIASMDPHN 353  
 QY 368 -----AVEGGLIGLTAFAFMILLVTAVTAAROVSRRLRRDRNPOAFMLASLAGLMLCHG 422  
 DB 354 ELLYMGVREGGLIPVG---LTLAACFCALRIYAARKGR-----MAMLLAEVYITLHA 403  
 QY 423 LEDVLYRPEASTLMLICIGATISFMOPO 451  
 DB 404 QLEYPFYH---SAHMITF-ITLIYWDQ 428

RESULT 13  
 O9AAUS PRELIMINARY; PRT; 440 AA.  
 AC O9AAUS;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)

DE HYPOTHETICAL PROTEIN CC0499.  
 GN CC0499.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=69394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nielsen W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka T., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouiri H., Shetty J., Berry K.,  
 RA Uitterback T., Iran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL: AE005722; AAK22486.1; -  
 DR TIGR: CC0499; -  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 440 AA; 47613 MW; D90CFDA3B6C5C5F CRC64;

Query Match 6.7%; Score 161.5; DB 2; Length 440;  
 Best Local Similarity 26.2%; Pred. No. 0.11;  
 Matches 95; Conservative 47; Mismatches 143; Indels 77; Gaps 17;

QY 70 LAAIAVWALLSLTIDLRQATPIHMLVILWGVDAATGLSP---VRAALVGLAKITL 126  
 DB 94 LAIVLRSMILLGLST-----AVCSVLMSIQPDVTIRLVAVVFTTLTG 137  
 QY 127 YLVFALAAVLRNRLSLFS--VVYITSLFVSVGLNOMIYGEELATWDRNSVAD 184  
 DB 138 VLL-----AERPEMKTLEVLASTAAVVTLSE--VPELFMPAVGVMSI-----D 180  
 QY 185 FTSRYSVSLGNPNLLA-AVLPPTTFSPAATG-----VMGWLPKLLAIAATGASSICL 237  
 DB 181 PFGANRGVYGRFNOGMWMSLATPFLACALANPQRRRLMGF-----AGALLLI 231  
 QY 238 ILTVSRGMLGFPVAMIFWALLGLYWFQRLPAPRRMLFPVVLGGLVAVLLV-AVIGLE 296  
 DB 232 VLSKTKIALVSGIAGLAMPILALCRIPALGT-----TALLAANSVALVGAVALA 284  
 QY 297 PLRVAVLSIFVGRDSSNNFRINWLAIVOMIQDRPWLGIGPGN-----TAFNLVYPLYQ 351  
 DB 285 PELVEEL---IGR-DATFTGRTLIMDAISROIQRPLTGYGYGAVWMDLISGMPVAMISN 340  
 QY 352 QARFALSAVSVPLEVAVNEGGLIGLTAFAMLLLVTAVTAAROVSRRLRRDRNPOAFMLAS 411  
 DB 341 DQGEFTFSAHNTVLGVLELIGVTA--AWMLLVGAM-LKGLSRL---ATAPAYFFLPF 394  
 QY 412 LA 413  
 DB 395 LA 396

RESULT 14  
 O06944 PRELIMINARY; PRT; 861 AA.  
 ID O06944;  
 AC O06944; O55995;  
 DT 01-JUN-1997 (Tremblrel. 04, Created)  
 DT 01-JUN-1997 (Tremblrel. 04, last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, last annotation update)  
 DE HYPOTHETICAL 96.7 KDA PROTEIN.  
 GN SL0737.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Suzuki T., Miyajima N.,

RA Sugiura M., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb  
 RT region from map positions 64k to 92k of the genome.";  
 RL DNA Res. 2:153-166(1995).  
 RN 12)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-97061201; Pubmed-8905231;  
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirosewa M., Sugiura M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Mureki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL: D64005; BAA10778.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 861 AA; 96682 MW; A064B98C2D9B6C59 CRC64;

Query Match 6.6%; Score 159; DB 2; Length 861;  
 Best Local Similarity 22.6%; Pred. No. 0.3;  
 Matches 112; Conservative 63; Mismatches 145; Indels 176; Gaps 30;

QY 48 GFLAVVYGSAPFVPS-ALGIG-----LAIAAYW-----ALLSLTDLR 88  
 DB 27 GLTLALLGLTTLNGLSLMGWFWVWQTWLTKLMOIGKRLISL----- 80  
 QY 89 QATPIHMLVLLWGYDALATGISPVA-----AAVLGAKLTLYLVFALA 134  
 DB 81 GLGFDMLVAL-ALAAMLSVGAEFPQARWGWIFCAFLAGLGRSWL----- 129  
 QY 135 ARVLNP--RLRSL-----FSVYVITSLFVSYYGL-----NQWLYG 169  
 DB 130 -KVQTPPLVTSINKLLTFQGYGFATVISLLMTQTLLPQWSAIAAREMGITKTFEG 188  
 QY 170 VEELETWDRNSVADFTSRVSYLGNPNMLAAVYPTAFSAAGVW-----RQMLPKL 224  
 DB 169 ILELONMP-----IGHQNVAGYLV--LILPLSTLYLNLNGKKRWMSV 232  
 QY 225 LAIAVYASSICLI--LTYSRGWLGFWAMLEWALLGLYWFQRLDPAWRMLFPVYL 281  
 DB 223 -----ALISGLIDFYTTSRGLLG-LATILLIYIIGILR--QLPRMWL----- 276  
 QY 282 GGLVAVLVAVYGL-EPLEVRVLSIFVGRDESSN-----NFRIVWLAVLQMDR 331  
 DB 277 -GL-GTLAAVAVGIFCTNDRLLTSFTGIMGQAGAFAYRLINEIG-W-----RMGSAH 329  
 QY 332 PWLGIGPNTAFNLVYPLY-----QQAFTALSAVSPLEAVGGLGLTAFAMLLV 385  
 DB 330 PWTGTGIGLNP--LOYOLYRPWAGRESEF-IYOLHSTPAOLFALGV-----WGILI 379  
 QY 386 TAVTAVROYSLRDRNDP-----AFWMAASLAGLGMCHGLPDTVL 428  
 DB 380 PSLLIITGLIMQILRSILKONIYVTRKNDVDSNVKILIMTILS-----ALLAYGVTSLTD 434  
 QY 429 YRPEASTLMWLCIGAI 444  
 DB 435 YQLDN-----VCISGV 445

RESULT 15  
 Q9XC97 PRELIMINARY; PRT; 360 AA.  
 AC Q9XC97;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
 DE PUTATIVE LIPID CORE SURFACE POLYMER LIGASE WABL.  
 GN WABL.  
 OS Klebsiella pneumoniae.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Klebsiella.  
 OX NCBI\_Taxid-573;  
 RN 11)  
 RP SEQUENCE FROM N.A.  
 RA Regue M., Climent N., Abitia N., Pique N., Coderech N., Merino S.,  
 RA Aguilar A., Tomas J.M.;  
 RT "Genetic characterization of the *Klebsiella pneumoniae* C3 waa gene  
 RT cluster involved in core lipopolysaccharide biosynthesis.";  
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF146532; AAD37765.1; -;  
 KW Ligase.  
 SQ SEQUENCE 360 AA; 40146 MW; 523EB6CE2139930 CRC64;

Query Match 6.3%; Score 154; DB 2; Length 360;  
 Best Local Similarity 24.9%; Pred. No. 0.24;  
 Matches 98; Conservative 54; Mismatches 138; Indels 104; Gaps 24;

QY 77 WALSLTDLIDLRQATPIHMLVLLWYG-----VDALATGISPVAALVGIKLTLYLVF 131  
 DB 40 WTLA-----AFETVAFSNWGHTPQHID-----SPTRHGY-----LTGYLLIM 80  
 QY 132 ALAARVLRNRLSLFSVYVITSLFVSYYGLNQWLYGVPELATWDRNSVADFTSRVYS 191  
 DB 81 TM--LIRDGRTRLMLAVVGG--ITVLSLRTL-----IDHTLV--LTERAVS 123  
 QY 192 YLGNPNLLAAYLVPTAFSA--ALGV-----NRGMLPKLIAIATGASSL 235  
 DB 124 -PENPG-----PTNVIDLAGYCGIGITILCGMLLEKASHWL-YLPVYIML-----V 167  
 QY 236 CLITTSRGWLGFWAMLEWALLGLYWFQRLDPAWRMLFPVYLGLVAVLVAVLGL 295  
 DB 168 MLITQSRGPILALV--LAVGCTLHLVETR-----RNLIIAALAVLVALLV--M 215  
 QY 296 EPLRVRVLSIFVGRD--SSNNFRIVWLAVLQMDRPLGIGPNTAFNLVYPLYQA 353  
 DB 216 TPVQDMLLRP--EELGTOSGLRLSTIMHTLSOMASQPLVNG--FSYELDFINSGE 269  
 QY 354 RFTALSAVSPLEAVGGLGLTAFAMLLIYAVTAVROYSLRDRNDPAWFLMASIA 413  
 DB 270 HIT--THSVYMGALLKAGIVG-----LLLLLAIACGLMQAMRKHHSTSFSLALFY 321  
 QY 414 GLAGMLGHGLFDVLYLRP-EASTLMWLCIGAIAS 446  
 DB 322 ALVFMAAQGMF--IISNPRTWVLFWLPILGIALS 353

Search completed: January 10, 2002, 23:10:21  
 Job time: 373 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:36:02 ; Search time 108.09 seconds

(without alignments)  
11135.948 Million cell updates/sec

Title: US-09-887-038-2

Perfect score: 1404  
Sequence: 1 atgactgtctggcaactct.....attcagagaaaaatgtag 1404

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.4	3.4	810	AAH52001	Mycobacterium tube
2	47.4	3.4	2232	AAO78915	Mycobacterium bovi
3	47.4	3.4	2232	AAO75519	Mycobacterium bovi
4	47.4	3.4	2232	AAAT9224	M. bovis p55 opero
5	47.4	3.4	2609	AAH02108	Mycobacterium tube
6	47.4	3.4	3120	AAO75518	M. tuberculosis in
7	47.4	3.4	3120	AAO78914	M. tuberculosis in
8	47.4	3.4	3120	AAV92223	M. tuberculosis in
9	47.4	3.4	3120	AAV20433	Mycobacterium tube
10	47.4	3.4	3120	AAK21816	M. tuberculosis in
11	47.4	3.4	3120	AAV68212	Inha gene nucleoti

C	12	44.4	3.2	10732	21	AAI0594	Gene encoding a su
	13	39.6	2.8	1723	18	AAE92220	M. bovis INH resis
	14	39.6	2.8	1723	18	AAE92219	M. tuberculosis IN
	15	39.2	2.8	58857	21	AAH58471	Nucleotide sequenc
C	16	37.8	2.7	6741	21	AAAI0595	Gene encoding a su
	17	36.8	2.6	2631	20	AAH34135	Mycobacterium spec
C	18	36.2	2.6	1450	12	AAO13408	Mycobacterium spec
	19	36.2	2.6	1450	12	AAO13408	Pseudorabies virus
C	20	35.8	2.5	1091	22	AAI61038	Human polynucleoti
	21	35.8	2.5	1124	22	AAI59252	Human polynucleoti
C	22	35.8	2.5	1124	22	AAI59252	Human polynucleoti
	23	35.6	2.5	10211	19	AAE62152	Murine delta-relat
C	24	35.6	2.5	117213	19	AAE62176	HSV-2 strain SB5 C
	25	35.2	2.5	390	19	AAV23486	HSV-2 strain SB5 C
	26	35.2	2.5	390	21	AAI13900	Pseudomonas Xcpv s
C	27	35.2	2.5	434	21	AAH78189	Pseudomonas alcali
	28	35.2	2.5	434	22	AAI28927	cDNA encoding huma
C	29	35.2	2.5	3978	21	AAE5785	Colon tumour relat
	30	35.2	2.5	4100	20	AAE32019	Type I polyketide
	31	35.2	2.5	4100	22	AAE90076	Human METH1 relate
	32	35.2	2.5	12249	21	AAE55840	L13855 cDNA clone.
	33	35.2	2.5	17612	19	AAE23494	Complete Mitomycin
	34	35.2	2.5	17612	21	AAAI3905	Pseudomonas xpc, O
	35	35.2	2.5	17612	22	AAE30870	Pseudomonas alcali
	36	35.2	2.5	18331	21	AAE55857	Complete nucleotid
	37	35	2.5	30001	18	AAE61016	Total DNA sequence
	38	35	2.5	30001	20	AAE05110	S. aureofaciens DN
	39	35	2.5	114955	20	AAE53491	Human adenosine Al
C	40	34.8	2.5	43280	18	AAE80413	Human adenosine Al
	41	34.6	2.5	1450	12	AAO13227	tylactone synthase
C	42	34.6	2.5	3058	18	AAE77137	PRV glycoprotein g
	43	34.6	2.5	3132	22	AAE06335	Single chain antiq
C	44	34.6	2.5	3646	21	AAE36714	DNA encoding human
	45	34.2	2.4	678	22	AAH34854	Nucleotide sequenc
							Human colon cancer

#### ALIGNMENTS

RESULT	1
ID	AAH52001 standard; DNA: 810 BP.
XX	AAH52001;
AC	AAH52001;
XX	04-SEP-2001 (first entry)
DT	
XX	
DE	Mycobacterium tuberculosis potential drug target gene SEQ ID 55.
XX	
KW	Drug target; growth; organism viability; characterisation; ds.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	WO200135317-A1.
XX	
PD	17-MAY-2001.
XX	
PF	13-NOV-2000; 2000WO-US31152.
XX	
PR	12-NOV-1999; 99US-0165086.
PR	12-NOV-1999; 99US-0165124.
PR	01-FEB-2000; 2000US-0179531.
XX	
PA	(REGC ) UNIV CALIFORNIA.
XX	
PI	Eisenberg D, Rotstein SH, Marcotte EM;
XX	
DR	WPI: 2001-329193/34.
XX	P-PSDB; AAG81150.
PT	Identifying nucleotide or polypeptide sequence for use as drug target,
PT	involves providing algorithm that analyzes a functional relationship
PT	between nucleotide or polypeptide sequences, and comparing the

```
PT sequences
XX
PS Disclosure; Page 85; 207pp; English.
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX
SQ Sequence 810 BP; 143 A; 249 C; 281 G; 137 T; 0 other;

Query Match          3.4%; Score 47.4; DB 22; Length 810;
Best Local Similarity 54.2%; Pred. No. 0.0035;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagcgactacccgcgtggaagtcgagtgtagggcgactactggcttga 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 56 cgtcgtcgcgttcacatgcacggttagccagagagagagcgccagctgtgtctca 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 cggccttcgttgctgtgctgtgacgagcggtgacgagcggtgagccgac 1189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 cgggttcgacccggtcggtggtatcagcgcacccggtcggtcgcaaaagccc 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1190 tgcggcgacatcgcaatcccaagccttltgtgtagtgctagctggccggttgg 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 176 cgctgtcgaactcgacgtgcaaaagagagaggaactgtgcagcttggccggtgg 232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
AAQ78915
ID AAQ78915 standard; DNA; 2232 BP.
XX
AC AAQ78915;
XX
DT 03-JUL-1995 (first entry)
XX
DE Mycobacterium bovis ps5 operon.
XX
KW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; ps5 gene;
KM vaccine; mycobacteria; ds.
XX
OS Mycobacterium bovis.
XX
FH Key Location/Qualifiers
FT CDS 494..1237
    /*tag= a
    /label= ORF1
    /tag= b
    /label= ORF2
    /note= "inhA gene"
XX
XX
XX WO9426312-A.
XX
XX
XX PD 24-NOV-1994.
XX
XX
XX PF 12-MAY-1994; 94WO-US05344.
XX
XX
XX PR 13-MAY-1993; 93NZ-0247620.
XX
XX PR 14-MAY-1993; 93US-0062409.
XX
XX PR 31-MAR-1994; 94US-0221742.
XX
XX
XX (BANE/) BANERJEE A.
```

```
PA (COLL/) COLLINS D M.
PA (DLS/) DE LISIE G W.
PA (JACO/) JACOBS W R.
PA (WILS/) WILSON T M.
PA (AGRE-) AGRESEARCH.
PA (COLL/) COLLINS D.
PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Banerjee A, Collins DM, De LISIE GW, Jacobs WR;
PI Wilson TM, Collins D;
XX
XX WPT: 1995-006366/01.
XX
XX P-PSDB; AAR66291; AAR66292.
XX
XX Gene target for isonicotinic acid hydrazide - used to develop
XX prods for diagnosis, treatment, prevention and studies involving
XX mycobacterial infections
XX
XX
XX PS Disclosure; Fig. 9A-9D; 76pp; English.
XX
XX
XX A cosmid library from M. bovis G4/100 was prepared in shuttle
XX vector pYU818 and transformed into Mycobacterium smegmatis ms2155.
XX The smallest plasmid obtained which conferred resistance to INH,
XX the target of action for isoniazid, was designated ps5. The
XX sequence of ps5 was determined (AAQ78915), revealing 2 large open
XX reading frames, the inhA gene comprising ORF2.
XX
XX
XX SQ Sequence 2232 BP; 411 A; 696 C; 752 G; 373 T; 0 other;
```

```
Query Match          3.4%; Score 47.4; DB 16; Length 2232;
Best Local Similarity 54.2%; Pred. No. 0.0049;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagcgactacccgcgtggaagtcgagtgtagggcgactactggcttga 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1311 cgtcgtcgcgttccacatgcacggttagccagagagagcgccagctgtgtctca 1370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 cggccttcgttgctgtgctgtgacgagcggtgacgagcggtgagccgac 1189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1371 cgggttcgacccggtcggtggtatcagcgcacccggtcggtcgcaaaagccc 1430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1190 tgcggcgacatcgcaatcccaagccttltgtgtagtgctagctggccggttgg 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1431 cgctgtcgaactcgacgtgcaaaagagaggaactgtgcagcttggccggtgg 1487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
AAQ75519
ID AAQ75519 standard; DNA; 2232 BP.
XX
XX
XX AAQ75519;
XX
XX
XX DT 03-JUL-1995 (first entry)
XX
XX DE Mycobacterium bovis ps5 operon.
XX
XX
XX KW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine;
KM ps5; ss.
XX
XX OS Mycobacterium bovis G4/100.
XX
XX
XX FH Key Location/Qualifiers
FT CDS 494..1234
    /*tag= a
    /label= ORF1
    /tag= b
    /label= ORF2
    /note= "inhA gene"
XX
XX
XX WO9426765-A.
```

FT	CDS	1256..2065
FT		/+leg= b
PN	US5686590-A.	
XX		
PD	11-NOV-1997.	
XX		
PF	12-MAY-1994;	94US-0241766.
XX		
PR	12-MAY-1994;	94US-0241766.
PR	14-MAY-1993;	93US-0062409.
XX		
PA	(AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.	
XX		
PI	Banerjee A, Collins DM, De lisle GW, Jacobs WR;	
PI	Wilson TM;	
XX		
DR	WPI: 1997-558202/51.	
DR	P-PSDB; AAM40806, AAM40809.	
XX		
PT	Nucleic acids encoding mycobacterial InhA proteins - associated with	
PT	isoniazid susceptibility and resistance	
XX		
PS	Disclosure: Column 39-44; 55pp; English.	
XX		

This sequence represents the coding sequence for a mycobacterial InhA protein. The encoded protein is involved in mycolic acid biosynthesis. The nucleic acids are useful in assessing the susceptibility of various strains of the *M. tuberculosis* complex to isoniazid (INH) type antibiotics, as decoys and as antisense oligonucleotides to prevent the expression of polypeptides associated with isoniazid (also referred to as isonicotinic acid hydrazide) resistance, and for the expression of the proteins, which may also have use in immunoassays for the detection of INH resistant strains. In the determination of whether an INH type

Sequence 2232 BP; 411 A; 696 C; 751 G; 374 T; 0 other;

Query Match	3.4%;	Score 47.4;	DB 18;	Length 2232;
Best Local Similarity	54.2%;	Pred. No. 0.0049;		
Matches	96;	Conservative	0;	Mismatches 81; Indels 0; Gaps
QY	1070	cgttgcgagcctactccgcgcgcgttcgaagtgcggtttgaggcgcgactactggcgtlta	1129	
Db	1311	cgttcgatcgcgtttcatcaatcgtcacgcgggttagccacagaagaaagcgcccacagcttgtgtctca	1370	
OY	1130	cggcgcttccttgccgcgcgttcgttcaacgcgcggttgacgcgcltgsgcaggttgagccgac	1189	
Db	1371	ccgggttcgaccgcgcgtcggcgcttgattcagcagcatalcacccaacccgctgcgcggcacaagggcc	1430	

Db 1431 cgcctcgtcgaactcgacgctgcacaagaacgagcaactctggccagctcttggcccgcgcggg 1487

ID	AAH02108 standard; DNA; 2609 BP.
XX	AAH02108;
XX	24-JUL-2001 (first entry)
XX	Mycobacterium tuberculosis nucleotide sequence SEQ ID NO:2101.
XX	
XX	

KW vaccine; primer; ds.  
KW catalytic subunit of proton-translocating ATPase; antimicrobial;  
KW translation elongation factor G; RecA recombinase; resistance;  
KW intracellular; diagnosis; translation elongation factor 1 $\alpha$ ; coat;  
KW

```

XX OS Mycobacterium tuberculosis.
XX PN WO200123604-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000MO-CA01150.
XX PR 28-SEP-1999; 99CA-2283458.
XX PR 19-MAY-2000; 2000CA-2307010.
XX PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
XX PI Picard FJ, Roy PH;
XX DR WPI; 2001-245006/25.
XX PS Nucleic acid sequences are used to generate universal probes and
XX PS primers which can be used to identify and detect the presence of algal,
XX PS archaeal, bacterial, fungal and parasitcal species in a test sample.
XX PS Disclosure; Page 1488-1489; 1580pp; English.
XX PS
XX CC The present invention describes a method for generating a repertory of
XX CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes
XX CC and/or primers are derived. The method comprises amplifying the nucleic
XX CC acids of determined algal, archaeal, bacterial, fungal and parasitcal
XX CC species with a combination of defined primer pairs. The method can be
XX CC used for producing probes and/or primers for detecting one or more
XX CC related microorganisms e.g. algae, archaea, bacteria, fungi and
XX CC parasites, for universal detection and for specific and ubiquitous
XX CC detection and identification of an algal, archaeal, bacterial, fungal
XX CC and parasitcal species, genus, family and group. A nucleic acid (1)
XX CC obtained using the method of the invention can be used for the universal
XX CC detection of any bacterium, fungus or parasite in a sample and for the
XX CC detection of at least one antimicrobial agent resistance gene or at
XX CC least one toxin gene. hexa nucleic acids are used for the specific and
XX CC ubiquitous detection and for identification of Streptococcus pneumoniae.
XX CC (1) can be used to design a therapeutic agent which is effective against
XX CC microorganisms. Microbial species or genus or family or phylum or group
XX CC which can be detected include Abiotropha adiacens, Bordetella sp.,
XX CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
XX CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
XX CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
XX CC provides faster results than substrate specificity tests as results can
XX CC be determined in an hour and improved accuracy is also achieved.
XX CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
XX CC which are given in the exemplification of the present invention.
XX SQ
Sequence 2609 BP; 471 A; 820 C; 875 G; 443 T; 0 other;

Query Match 3.4%; Score 47.4; DB 22; Length 2609;
Best Local Similarity 54.2%; Pred. No. 0.0052;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 1070 cgttgagcgccactccgctccgctggaagtgcggttgagggcggaactacgtgggttga 1129
DB 1855 cgtcagtcgcgtttacatcagcagggtagccagagagagggcgccctggtgtccta 1914
QY 1130 cgccttcgcttggtcgtcgtgtgtacagcgcggtgacgagcggtgagcgagcgcgc 1189
DB 1915 ccgggttcacgcggcgctgcatcagcagcagcagcagcgccgagcaaaagccc 1974
QY 1190 ttgcgcgcgcatcgaaatccccaagcctttgttgatgtagctgagcgcggtttgg 1246
DB 1975 cgctgtcgaactcgacgctgcaaaacagagagagcactgtgcagcttgccggcgcg 2031

```

RESULT 6  
AA075518

```

ID AA075518 standard; DNA; 3120 BP.
XX AC AA075518;
XX AC
XX DT 03-JUL-1995 (first entry)
XX DE M. tuberculosis inhA operon.
XX DE
XX KW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine; ss.
XX OS Mycobacterium tuberculosis.
XX PN WO9426765-A.
XX PD 24-NOV-1994.
XX PF 13-MAY-1994; 94MO-US05398.
XX PR 13-MAY-1993; 93NZ-0247620.
XX PR 14-MAY-1993; 93US-0062409.
XX PR 31-MAR-1994; 94US-0221742.
XX PA (AGRE-) AGRESEARCH.
XX PA (BANE/) BANERJEE A.
XX PA (COL/) COLLINS D.
XX PA (JACO/) JACOBS W R.
XX PA (YESH/) UNIT YESHIVA EINSTEIN COLLEGE.
XX PA (WILS/) WILSON T M.
XX PI Banerjee A, Collins D, de LISLE GW, Jacobs WR, Wilson TM;
XX PI WPI; 1995-006691/01.
XX DR P-PSDB; AAR66901.
XX PS Polynucleotide(s) determining mycobacterial resistance to
XX PS isoniazid - useful in diagnosis, treatment and prevention of
XX PS mycobacterial infection, e.g. tuberculosis.
XX PS Disclosure; Page 42-44; 104pp; English.
XX PS
XX CC The gene from Mycobacterium tuberculosis encoding inhA (AAR66901),
XX CC the target of action for isoniazid, was identified, isolated,
XX CC cloned and sequenced (AA075518). Mutant inhA genes have been used
XX CC for recombinant vaccine development.
XX SQ
Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match 3.4%; Score 47.4; DB 16; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 1070 cgttgagcgccactccgctccgctggaagtgcggttgagggcggaactacgtgggttga 1129
DB 1041 cgtcagtcgcgtttacatcagcagggtagccagagagagggcgccctggtgtccta 1100
QY 1130 cgccttcgcttggtcgtcgtgtgtacagcgcggtgacgagcggtgagcgagcgcgc 1189
DB 1101 ccgggttcacgcggcgctgcatcagcagcagcagcagcgccgagcaaaagccc 1160
QY 1190 ttgcgcgcgcatcgaaatccccaagcctttgttgatgtagctgagcgcggtttgg 1246
DB 1161 cgctgtcgaactcgacgctgcaaaacagagagagcactgtgcagcttgccggcgcg 1217

```

RESULT 7  
AA078914  
ID AA078914 standard; DNA; 3120 BP.  
XX AC AA078914;  
XX AC  
XX DT 03-JUL-1995 (first entry)  
XX

```
DE M. tuberculosis inhA gene.
XX
KM Isoniazid; isonicotinic acid hydrazide; INH; inhA gene;
KM vaccine; mycobacteria; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN W09426312-A.
XX
PD 24-NOV-1994.
XX
PF 12-MAY-1994; 94WO-US05344.
XX
PR 13-MAY-1993; 93N2-0247620.
PR 14-MAY-1993; 93US-0062409.
PR 31-MAR-1994; 94US-0221742.
XX
PA (BANE/) BANERJEE A.
PA (COLL/) COLLINS D M.
PA (DLIS/) DE LISLE G W.
PA (JACO/) JACOBS W R.
PA (WILS/) WILSON T M.
PA (AGRE-) AGRESEARCH.
PA (COLL/) COLLINS D.
PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Banerjee A, Collins DM, De Lisle GW, Jacobs WR;
PI Wilson TM, Collins D;
XX
DR WPI; 1995-006366/01.
DR P-PSDB; AAR66290.
XX
PT Gene target for isonicotinic acid hydrazide - used to develop
PT prods for diagnosis, treatment, prevention and studies involving
PT mycobacterial infections
XX
PS Disclosure; Fig. 8A1-8C2; 76pp; English.
XX
CC Genes from Mycobacterium smegmatis, Mycobacterium tuberculosis and
CC Mycobacterium bovis that encode inhA, the target of action for
CC isoniazid, were identified, isolated and cloned. Sequences of
CC the 3 genes are given in AAQ78913-15, and encoded amino acids in
CC AAR66289-91. Mutant genes have been used in recombinant vaccine
CC development.
XX
SQ Sequence 3120 BP; 562 A; 1006 C; 1040 G; 512 T; 0 other;

Query Match 3.4%; Score 47.4; DB 16; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttgagcgctactcgtcccgcttggaagtcgaggttgagcgagacttggttga 1129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1041 cgtcgatcgcttcaacatcgacggtgtagccaggaagcgagcagctgtgtctca 1100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 cggccttcgctgtgctgtgtgtcagcgagcggtgagcgaggttgagcgagc 1189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1101 ccgggttcgacggctgctgggtgtatcagcgatcacccgacggcttgcggaaggcc 1160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1190 tgcggcgagatcgcaatcccaagccttggltgatgctagcttgccggtttg 1246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1161 cgcctcgcgaactcgacgtgcaaaacgagagcacccttgccagcttggccgcggg 1217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AAV9223
ID AAV9223 standard; DNA; 3120 BP.
XX
AC AAV9223;
XX
AT 01-APR-1998 (first entry)
XX
DE
```

```
DE M. tuberculosis inhA gene.
XX
KM INH resistance; inhA gene; isonicotinic acid hydrazide; mycobacteria;
KM mycolic acid biosynthesis; antibiotic susceptibility; infection therapy;
KM INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;
KM antibiotic resistant strain; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US5686590-A.
XX
PD 11-NOV-1997.
XX
PF 12-MAY-1994; 94US-0241766.
XX
PR 12-MAY-1994; 94US-0241766.
PR 14-MAY-1993; 93US-0062409.
XX
PA (AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.
XX
PI Banerjee A, Collins DM, De Lisle GW, Jacobs WR;
PI Wilson TM;
XX
DR WPI; 1997-558202/51.
XX
PT Nucleic acids encoding mycobacterial inhA proteins - associated with
PT isoniazid susceptibility and resistance
XX
PS Disclosure; Column 37-40; 55pp; English.
XX
CC This sequence represents the coding sequence for a mycobacterial inhA
CC protein. The encoded protein is involved in mycolic acid biosynthesis.
CC The nucleic acids are useful in assessing the susceptibility of various
CC strains of the M. tuberculosis complex to isoniazid (INH) type
CC antibiotics, as decoys and as antisense oligonucleotides to prevent the
CC expression of polypeptides associated with isoniazid (also referred to as
CC isonicotinic acid hydrazide) resistance, and for the expression of the
CC proteins, which may also have use in immunoassays for the detection of
CC INH resistant strains, in the determination of whether an INH type
CC antibiotic may be effective against tuberculosis, and in the treatment of
CC individuals for infection with these strains.
XX
SQ Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match 3.4%; Score 47.4; DB 18; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttgagcgctactcgtcccgcttggaagtcgaggttgagcgagacttggttga 1129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1041 cgtcgatcgcttcaacatcgacggtgtagccaggaagcgagcagctgtgtctca 1100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1130 cggccttcgctgtgctgtgtgtcagcgagcggtgagcgaggttgagcgagc 1189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1101 ccgggttcgacggctgctgggtgtatcagcgatcacccgacggcttgcggaaggcc 1160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1190 tgcggcgagatcgcaatcccaagccttggltgatgctagcttgccggtttg 1246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1161 cgcctcgcgaactcgacgtgcaaaacgagagcacccttgccagcttggccgcggg 1217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AAV20433
ID AAV20433 standard; DNA; 3120 BP.
XX
AC AAV20433;
XX
AT 15-JUN-1998 (first entry)
XX
DE Mycobacterium tuberculosis inhA gene.
XX
KM Mycobacterium tuberculosis inhA gene.
XX
DE
```

```

KM X-ray crystallography; three dimensional structure; crystallised; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN US5702935-A.
XX
PD 30-DEC-1997.
XX
PF 28-APR-1994; 94US-0234011.
XX
PR 28-APR-1994; 94US-0234011.
XX
PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Blanchard J, Jacobs WR, Sacchettini J;
XX
DR WPI; 1998-076420/07.
XX
PT Crystalline forms of InhA enzyme - to help in developing
XX
PS anti-tuberculosis agents or herbicides
XX
PS Example 1; Column 15-18; 22pp; English.
XX
CC The present sequence represents the InhA gene from Mycobacterium
CC tuberculosis. The present invention describes: (1) a crystallised
CC InhA enzyme in the form of a plate having the space group C2 and the
CC unit cell constants a = 101.1 Angstrom, b = 83.4 Angstrom, c = 192.9
CC Angstrom, beta = 95 deg. and alpha = gamma = 90 deg.; (2) a
CC crystallised InhA enzyme in the form of a parallelepiped having the
CC space group P21 and the unit cell constants a = 65.0 Angstrom, b =
CC 116.0 Angstrom, c = 65.0 Angstrom, beta = 97.8 deg. and alpha =
CC gamma = 90 deg.; (3) a crystallised InhA enzyme in the form of a
CC hexagon having the space group P622 and the unit cell constants a =
CC b = 100.14 Angstrom, c = 140.4 Angstrom, alpha = beta = 90 deg. and
CC gamma = 120 deg. A knowledge of the crystal structure of InhA (which
CC is the target of isoniazid in Mycobacterium tuberculosis) will help in
CC the development of bactericides or herbicides capable of inhibiting
CC InhA activity.
XX
SQ Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match 3.4%; Score 47.4; DB 19; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagagcgctactccgctccgctggaagtcggttgaggcggaactacgggttga 1129
DB 1041 cgtcagtcgcttcacatcgacgaggtatgcccagagacagggcgcccgctggtctca 1100
QY 1130 cggccttcgctgctgctgctgctgcaacgagcggtgacggcgagtgagccgac 1189
DB 1101 cggggttcgacgagctgctgctgctgctgcaacgagcggtgacggcgagtgagcc 1160
QY 1190 tcgcgagcgatcgcaatcccaagcctttggttgatgagctgagctggcggtttg 1246
DB 1161 cgcgtcgcgaactcgacgtgcaaaagagagacactggtgcggtgcgcggcg 1217

RESULT 10
AAK21816
ID AAK21816 standard; DNA; 3120 BP.
XX
AC AAK21816;
XX
DT 14-MAY-1999 (first entry)
XX
DE M. tuberculosis inhA gene.
XX
KW InhA; enzyme; inhibitor; infection; therapy; ss.
XX
OS Mycobacterium tuberculosis.
XX

```

```

PN US5882878-A.
XX
PD 16-MAR-1999.
XX
PF 21-AUG-1996; 96US-0701062.
XX
PR 16-SEP-1994; 94US-0307376.
XX
PR 28-APR-1994; 94US-0234011.
XX
PR 16-JUN-1995; 95US-0491146.
XX
PR 21-AUG-1996; 96US-0701062.
XX
PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
XX
PI Sacchettini J;
XX
DR WPI; 1999-214062/18.
XX
PT Identifying inhibitors of InhA enzyme - by identifying compound
XX
PS that fits the catalytic active site of the crystallised enzyme,
XX
PS useful for treating tuberculosis
XX
PS Disclosure; Column 3-8; 9pp; English.
XX
CC This sequence represents the Mycobacterium tuberculosis inhA coding
CC sequence. The invention relates to a method for identifying inhibitors of
CC InhA enzyme comprises identifying compounds which fit the catalytic
CC active site of the crystallised enzyme. Inhibitors of the activity of the
CC InhA enzyme can be used for treating Mycobacteria tuberculosis infection.
XX
SQ Sequence 3120 BP; 566 A; 995 C; 1037 G; 522 T; 0 other;

Query Match 3.4%; Score 47.4; DB 20; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.0056;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagagcgctactccgctccgctggaagtcggttgaggcggaactacgggttga 1129
DB 1041 cgtcagtcgcttcacatcgacgaggtatgcccagagacagggcgcccgctggtctca 1100
QY 1130 cggccttcgctgctgctgctgctgcaacgagcggtgacggcgagtgagccgac 1189
DB 1101 cggggttcgacgagctgctgctgctgctgcaacgagcggtgacggcgagtgagcc 1160
QY 1190 tcgcgagcgatcgcaatcccaagcctttggttgatgagctgagctggcggtttg 1246
DB 1161 cgcgtcgcgaactcgacgtgcaaaagagagacactggtgcggtgcgcggcg 1217

RESULT 11
AAV68212
ID AAV68212 standard; DNA; 3120 BP.
XX
AC AAV68212;
XX
DT 25-JAN-1999 (first entry)
XX
DE inhA gene nucleotide sequence.
XX
KW InhA enzyme; Mycobacterium tuberculosis strain H37 Rv; bactericide;
XX
KW enoyl reductase activity; mycolic acid production; herbicide; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN US5837480-A.
XX
PD 17-NOV-1998.
XX
PF 21-AUG-1996; 96US-0700306.
XX
PR 28-APR-1994; 94US-0234011.
XX
PR 21-AUG-1996; 96US-0700306.
XX

```





Dd	9878	YDAGSUDBIYRCRCCTTYYISSTDISTDSTDNSTTBSCDCTTTTBRSNSTSISSTIYRC	9819
Qy	925	cgtgaagcaagcaagcaacaactcccgatcaatgctgctgctgcgctgctgcaatgatt	984
Dd	9818	SDYDABSDNSNTMCCYDASRTBIBSTBNCYARCIYBYDARCSDSTYSARGDANSTRYS	9759
Qy	985	caagtcgagcttcgtctggcatcgagcccgcgacataccgcttlaacctggttacc	1044
Dd	9768	SNYSSTYSDDSTYSKAYCAKSTTBIBCYDAYDACYDAYDMNCYSBSTYBYCSRRC	9699
Qy	1045	ctctcaacacagcgcgcttaccgacgctgtagcgactaccgctcccgctggaagtcgc	1103
Dd	9698	YVDAYSCGRYDARCADYCAYSNSRVCYDABTBSRSTYSYVNCYDATTBSRCTBYSTB	9639
Qy	1104	ggttgaaggcggaactactggtgcttgaagcgcttcgctgctgctgctgctgatacgcg	1163
Dd	9638	TBTTTRCKACTBDSSTAKNSTSYSTRCBYBSRSNGYCSRSRRCYSCYTDSST	9579
Qy	1164	gacgcgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1223
Dd	9578	CYSTIYSTTAYASCTTSNGSYDASRSYSTSRCTTSYSTSTDDYSDCSYSTTBNST	9519
Qy	1224	gatggctagcttgccggttctgcgcgaagatgctggatcaggtctgttatacgcgct	1283
Dd	9518	YSSDSCBYSBSSDRCSRSDSTCNCYSCDSRSYSTTYDACYDYDAKTBCTYVSDN	9459
Qy	1284	ctatcgaccgcgaagcgaatcgcctctgctgctctgataatggaacgcgcgcgcgcgc	1343
Dd	9458	SRCTYNSTYSRBSRBSRBSRSTCTBTTSRNGCCYDADANSTRYDAYDACYSYDAST	9399
Qy	1344	gcagcccaacctlcccaagcaactccctcca	1374
Dd	9398	TBYSCTBYSNSTYDAYSSRYSCYTCYTC	9368
RESULT 13			
AAT99220			
ID	AAT99220 standard; DNA; 1723 BP.		
xx			
AC	AAT99220;		
xx			
DT	01-APR-1998 (first entry)		
xx			
DE	M. bovis INH resistance gene.		
xx			
KW	INH resistance; INHase; isonicotinic acid hydrazide; mycobacteria;		
KW	mycolic acid biosynthesis; antibiotic susceptibility; infection therapy		
KW	INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;		
KW	antibiotic resistant strain; ss.		
xx			
OS	Mycobacterium bovis.		
xx			
PN	US5686590-A.		
xx			
PD	11-NOV-1997.		
xx			
PP	12-MAY-1994; 94US-0241766.		
xx			
PR	12-MAY-1994; 94US-0241766.		
xx			
PR	14-MAY-1993; 93US-0062409.		
xx			
PA	(AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.		
xx			
PI	Banerjee A, Collins DM, De Lisle GW, Jacobs WR,		
xx			
PI	Wilson TM;		
xx			
DR	WPI; 1997-558202/51.		
xx			
DR	P-PSDB; AAM40806.		
xx			
PT	Nucleic acids encoding mycobacterial INH proteins - associated with		
xx			
PT	isoniazid susceptibility and resistance		
xx			
PS	Disclosure; Column 21-24; 55pp; English.		

XX This sequence represents the coding sequence for a mycobacterial InhA  
CC protein. The encoded protein is involved in mycolic acid biosynthesis.  
CC The nucleic acids are useful in assessing the susceptibility of various  
CC strains of the *M. tuberculosis* complex to isoniazid (INH) type  
CC antibiotics, as decoys and as antisense oligonucleotides to prevent the  
CC expression of polypeptides associated with isoniazid (also referred to as  
CC isoniicotinic acid hydrazide) resistance, and for the expression of the  
CC proteins, which may also have use in immunoassays for the detection of  
CC INH resistant strains, in the determination of whether an INH type  
CC antibiotic may be effective against tuberculosis, and in the treatment of  
CC individuals for infection with these strains.  
XQ Sequence 1723 BP; 328 A; 530 C; 580 G; 285 T; 0 other;

Query Match	2.88;	Score 39.6;	DB 18;	Length 1723;
Best Local Similarity	55.18;	Pred. No. 0.58;		
Matches 98;	Conservative 0;	Mismatches 79;	Indels 1;	Gaps 1.

QY	1070	cgttcagacgcgtactacccgcgtccgcgtcgtgaagctcgcgtgtgttgagagcgcgcgaactacatcgtgcttga	1129
Db	968	cgttcagacgcgtgttcaacacacgcagcgtgtagccacagagcagagcgcgcgcacacgtgctgtca	1027
QY	1130	cgcgccttcgccttcgtcgtcgtcgtcgtg-ctcaacgcgcgtgtgaaagcgcgtgtgcgcgcaggtgtgaagccga	1188
Db	1028	cgcgtgttcagacgcgcgtcgtcgtcgtatctacagcgtatcaacgcgcgcgtcgtcgcgcgaagccgc	1087
QY	1189	ctgcgcgcgcgtatccaatcccaagcgtcttctgtgtatgagctaaccttcgcgcgcgttttg	1246
Db	1088	cgcgtcgtcgaactcagacgtgcaaaacgcagagcgcgcacacgtgcacagcttcgtccgcgcgcgtg	1145

RESULT	14
AAT99219	
ID	AAT99219 standard; DNA; 1723 BP.
XX	
AC	AAT99219;
XX	
DT	01-APR-1998 (first entry)
XX	
DE	M. tuberculosis INH resistance gene.
XX	
KW	INH resistance; inhA gene; isonicotinic acid hydrazide; mycobacteria;
KW	mycolic acid biosynthesis; antibiotic susceptibility; infection therapy;
KW	INH type antibiotic; M. tuberculosis complex; isoniazid; tuberculosis;
KW	antibiotic resistant strain; ss.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	US5686590-A.
XX	
PD	11-NOV-1997.
XX	
PF	12-MAY-1994; 94US-0241766.
XX	
PR	12-MAY-1994; 94US-0241766.
PR	14-MAY-1993; 93US-0062409.
XX	
PA	(AGRE-) AGRESEARCH NEW ZEALAND PASTORAL AGRIC.
XX	
PI	Banerjee A, Collins DM, De Lisle GW, Jacobs WR,
PI	Wilson TW;
XX	
DR	WPI; 1997-558202/51.
XX	
DR	P-PSDB; AAM40805.
XX	
PT	Nucleic acids encoding mycobacterial inhA proteins - associated with
PT	isoniazid susceptibility and resistance
XX	
PS	Claim 1; Column 19-22; 55pp: English.
XX	
CC	This sequence represents the coding sequence for a mycobacterial inhA





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:32:57 ; Search time 1447.17 Seconds

16005.058 Million cell updates/sec

Title: US-09-887-038-2

Perfect score: 1404

Sequence: 1 atgacgtctgcgcgaactc.....atccagcgcgaanaatgtag 1404

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pal:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_pl:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rtd:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1404	100.0	4957	1	SPU62616	U62616 <i>Synechococcus</i>
2	326.4	23.2	143051	1	D90911	D90911 <i>Synechocyst</i>
3	49.8	3.5	7218	6	166494	166494 Sequence 14
4	47.4	3.4	810	1	AF106077	AF106077 <i>Mycobacte</i>
5	47.4	3.4	832	1	MTU02492	MTU02492 <i>Mycobacte</i>
6	47.4	3.4	1789	1	MBU41388	MBU41388 <i>Mycobacte</i>
7	47.4	3.4	2232	6	173546	173546 Sequence 12
8	47.4	3.4	2609	6	AX111368	AX111368 Sequence
9	47.4	3.4	3120	6	126125	126125 Sequence 1
10	47.4	3.4	3120	6	173545	173545 Sequence 11
11	47.4	3.4	3120	6	186995	186995 Sequence 1
12	47.4	3.4	20113	1	AE007022	AE007022 <i>Mycobacte</i>
13	47.4	3.4	38300	1	MTY277	MTY277 <i>Mycobacte</i>
14	44.4	3.2	10732	6	E32986	E32986 Gene encodi
15	42.8	3.0	346294	1	AE002999	AE002999 <i>Mesorhizo</i>
16	41.8	3.0	20021	1	AE004730	AE004730 <i>Pseudomon</i>
17	41.6	3.0	2793	1	AF254803	AF254803 <i>Burkholde</i>
18	41.2	2.9	37890	2	AC092316	AC092316 <i>Homo sapi</i>
19	40.8	2.9	127488	2	AC022095	AC022095 <i>Homo sapi</i>
20	40.4	2.9	45335	1	SCC57A	AL136519 <i>Streptomy</i>
21	39.8	2.8	189333	1	RME603647	AL603647 <i>Rhizobium</i>
22	39.6	2.8	1723	6	173541	173541 Sequence 1
23	39.6	2.8	1723	6	173542	173542 Sequence 2
24	39.6	2.8	13569	1	AE004772	AE004772 <i>Pseudomon</i>
25	39.6	2.8	70089	2	AC027491	AC027491 <i>Homo sapi</i>
26	39.4	2.8	34766	1	SCC121	AL137166 <i>Streptomy</i>
27	39.4	2.8	137137	9	AC004928	AC004928 <i>Homo sapi</i>
28	39.2	2.8	1246	6	AX164174	AX164174 Sequence
29	39.2	2.8	9316	1	AF077869	AF077869 <i>Streptomy</i>
30	39.2	2.8	77457	1	AF210249	AF210249 <i>Streptomy</i>
31	39.2	2.8	168347	2	AC025336	AC025336 <i>Homo sapi</i>
32	38.8	2.8	5023	3	DMU11383	U11383 <i>Drosophila</i>
33	38.8	2.8	6655	3	DMOVO	X59772 <i>D.melanogas</i>
34	38.8	2.8	12991	1	AE007016	AE007016 <i>Mycobacte</i>
35	38.8	2.8	39150	1	MTY218A	MTY218A <i>Mycobacte</i>
36	38.6	2.7	13283	1	AE005006	AE005006 <i>Halobacte</i>
37	38.6	2.7	172154	2	AC087726	AC087726 <i>Chlamydom</i>
38	38.4	2.7	5186	1	RCAMCPAB	LA8927 <i>Rhodobacter</i>
39	38.4	2.7	10757	1	AE004561	AE004561 <i>Pseudomon</i>
40	38.4	2.7	10944	1	AE004771	AE004771 <i>Pseudomon</i>
41	38.2	2.7	1681	1	SC6310A	AL445343 <i>Streptomy</i>
42	38.2	2.7	11477	1	AE005892	AE005892 <i>Caulobact</i>
43	38	2.7	132712	2	AF004009	AF004009 <i>Oryza sat</i>
44	38	2.7	136555	2	AC068279	AC068279 <i>Homo sapi</i>
45	38	2.7	205061	2	AC069341	AC069341 <i>Homo sapi</i>

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE
1	SPU62616	<i>Synechococcus</i> PCC7942 putative protein (dc11) gene, partial cds, and putative proteins (dc12), (dc13), (dc14) and (dc13) genes, complete cds.	U62616	U62616.1	GI:1549372		<i>Synechococcus</i> sp. PCC 7942.	1 (bases 1 to 4957)	Bonfil,D.J., Lieman-Hurwitz,J., Ronen-Tarazi,M. and Kaplan,A.	Genomic region involved in ability of <i>Synechococcus</i> PCC 7942 to grow under low CO2 conditions	Unpublished	2 (bases 1 to 4957)	Bonfil,D.J., Lieman-Hurwitz,J., Ronen-Tarazi,M. and Kaplan,A.	Direct Submission

JOURNAL Submitted (01-JUL-1996) Plant Biology, Hebrew University, Givat  
Ram, Jerusalem 91904, Israel  
FEATURES Location/Qualifiers  
source 1..4957  
/organism="Synecchococcus sp. PCC 7942"  
/strain="PCC 7942"  
/db\_xref="taxon:1140"  
1..845  
/gene="dc11"  
1..845  
/gene="dc11"  
/note="similar to the C-terminus of a Synecchocystis  
PCC6803 putative protein encoded by GenBank Accession  
Number D64000 (cyanobase N. s110524)"  
/codon\_start=3  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAB08473.1"  
/db\_xref="GI:1549373"  
/translation="AMIEAIGVNPSPDDPOFILMVPFSAKIEDLLQGLDFAPPAVK  
GGLASGAGESAIAISLFQDQVNDVCIGALSNVOIAIVAGCGRPYPMHVAAE  
RNILRQLQTEDEPIAALQALDSYLRDLSPELRSLCYGLACNSQTYVQEDPIIRN  
LIGFDPRTGAVAIIGDIRVGRQLQHLHVDQTAUDLEKQLQWCQRQHATPPAALLF  
SCIGRGPYQQA NESQLIQHLYSELPLAGFCGEGIGPIAGSTYLIQYTSVALIS  
AKTH"  
misc\_feature 132..188  
/gene="dc11"  
/note="encodes a putative transmembrane region"  
misc\_feature 147..158  
/gene="dc11"  
/note="encodes a glycosaminoglycan attachment site"  
237..269  
/gene="dc11"  
/note="encodes a prokaryotic membrane lipoprotein lipid  
attachment site"  
411..473  
/gene="dc11"  
/note="encodes a putative transmembrane region"  
723..797  
/gene="dc11"  
/note="encodes a putative transmembrane region"  
misc\_feature complement(1095..1928)  
/gene="d133"  
/note="encodes a putative transmembrane region"  
complement(1095..1928)  
/gene="d133"  
/note="similar to Synecchocystis PCC6803 cyanobase N.  
s1r1895 and to E. coli and H. influenzae ICC proteins,  
Swiss-Prot Accession Numbers P36650 and P44685  
respectively; similar to a 12.8 KDa putative protein in  
the 16S rRNA gene region of Synecchococcus PCC6301,  
Swiss-Prot Accession Number P05677"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAB08474.1"  
/db\_xref="GI:1549374"  
/translation="MAMNRAFRFALLSDHLIGLPOTIMHHPHFILIECSIPAFROIL  
TELAIDLDFLLPGDILQNGERENHMOLOLROLPYVYVYVPPNHNHMAAASROIL  
NGVAFVELYDQFGTSDRPYRCITAPLDLDFANSTAFDAGQLRGKRDCKOLN  
WLOEELASTASQKRWVTHHNHLEHLPDQOQHPDQRTILEAEELRSITLEGSYSVI  
FTGHILQATSSQGALMEVHTGSLVSPHPRYLGTVPAPAEQMOLLESRRVRLSPD  
D"  
misc\_feature complement(1311..1322)  
/gene="d133"  
/note="encodes an endoplasmic reticulum targeting  
sequence"  
gene 1353..1748  
/gene="dc12"  
1353..1748  
/gene="dc12"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"

/protein\_id="AAB08475.1"  
/db\_xref="GI:1549375"  
/translation="MLLIIGMLOHIVNHNHPALASGSGSEFLLQPELRAPDSPMO  
LMARFKSDRQKQKQIEKCDRAIIVAKIAEAKVLLEDKLGHITGTAIARCGGHV  
VARHIDRIQGLAQLLQOPLWFPPLVIG"  
complement(1431..1442)  
/gene="d133"  
/note="encodes an endoplasmic reticulum targeting  
sequence"  
misc\_feature 1683..1745  
/gene="dc12"  
/note="encodes a putative transmembrane region"  
1998..2639  
/gene="dc13"  
1998..2639  
/gene="dc13"  
/note="cytoplasmic protein; similar to from Synecchocystis  
PCC6803 cyanobase N. s111300, to Escherichia coli YGGH  
gene product, Swiss-Prot Accession Number P32049, to  
Haemophilus influenzae YGGH gene product, Swiss-Prot  
Accession Number P44648 and to a region in the mica/mutY  
gene from Escherichia coli"  
/codon\_start=1  
/transl\_table=11  
/product="putative protein"  
/protein\_id="AAB08476.1"  
/db\_xref="GI:1549376"  
/translation="MARYRVKQHVNPISQKFOYVTPPMQOYVADCDRLPHLDIGCA  
NGRFLAMARQPEWNTYLGLEIREPLVDENALAREBELNLYHFSNALDLEPLR  
SLPTGILQKRSVSIQFPDPMFKRHQKRRVQPELVQALATAPLPGAEVFLQSDVLEVOA  
EMCEHFAEPREFQRTCIDWLPENLPVPTREIATVQKQDLPVYRALFIRPAD"  
2160..2180  
/gene="dc13"  
/note="encodes a tyrosine kinase phosphorylation site"  
2670..4073  
/gene="dc14"  
2670..4073  
/gene="dc14"  
/note="inactivation of this gene produced a  
high-CO2-requiring mutant; similar to Synecchocystis  
PCC6803 cyanobase N. s1r1515 and to Escherichia coli  
sodium/pantothenate symporter and a multidrug resistance  
protein, Swiss-Prot Accession Numbers P16256 and P21441  
respectively"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="AAB08477.1"  
/db\_xref="GI:1549377"  
/translation="MTVMQTLTEAHYQPOQMGHSFLHRLFGSLKAWRASSQLWVSE  
ALGGLFLLAVYGSAPFVPSALGGLAIAAYMALSLTDIDROATPIMHVLVWG  
VDALATGSPVRAAALVGLAKITGLYLFVLAARVLRNPRLSLPSVVTSLPVSIV  
YGLNQWITYGVELATWDRNSVADFTSRVSYLGNPNLAIYVPTTAPSAALIGYWR  
GWLKRLAIATGASSLCLLITITSKGGMLGFWAIVFWALLGLYWPQPLPAWRKRL  
FPVVLGGIVAVLVAIVGLPELRVRLVSLVGRDSSNNRINVMALVLMIDQPL  
GIGGNTAFNIVLYLQOARFTALASVYAEVASEGGLGALFAFLVLTAVTAVR  
QVSLRDRNRNQFWMLASLAGMIGHGIFDVIYLRPASTLMMLCIGALISFMQF  
QPSQQLPPEAEHSDKKM"  
2847..2918  
/gene="dc14"  
/note="encodes a putative transmembrane region"  
2952..3035  
/gene="dc14"  
3012..3074  
/gene="dc14"  
/note="encodes a putative transmembrane region"  
3081..3146  
/gene="dc14"  
/note="encodes a putative transmembrane region"  
3102..3158  
/gene="dc14"  
/note="encodes an ATP synthase c subunit signature"  
/note="encodes a putative transmembrane region"

```

misc_feature      3264..3320
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3400..3455
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3492..3554
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3783..3842
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3882..3941
                  /note="encodes a putative transmembrane region"
misc_feature      3972..4022
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
misc_feature      3972..4022
                  /gene="dc14"
                  /note="encodes a putative transmembrane region"
BASE COUNT      972 a 1418 c 1451 g 1116 t
ORIGIN
Query Match      100.0%; Score 1404; DB 1; Length 4957;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 atgactgtctgcaaacctgactcttgcacattacacaccccaacagtgaggccacagc 60
    |||||||
Db 2670 ATGACTGTCTGCAAACTCTACTCTTTGCGCATTAACCAACCCCAACATGGGGCCACAGC 2729

Oy 61 agttcttgcacacgctgttctgacagcctgagccttgccggccctccacagcagtgtg 120
    |||||||
Db -2730 AGTTCTTGTCATCGGCTCTTTGGGAGCCTGGAGCTTGGGGGCTCCAGCAGCTGTG 2789

Oy 121 gttgtgtctgagcagcagctgtgtgtctctgtctgtctgtctgtctgtctgtctgt 180
    |||||||
Db 2790 GTTGTGTCTGAGGCACTGGGTGGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2849

Oy 181 gtggccaaattccgcgcctgaggttggttggttggttggttggttggttggttggttc 240
    |||||||
Db 2850 GTGCCAATTCGCGCCTTAAGGCTTGGGGCTACCGCGAGTGGCGCTATTTGGGCTCGTC 2909

Oy 241 tgcgtgacagatatacgtatctgcgcaagcaaccccccaattcactgtgctgtctgtctac 300
    |||||||
Db 2910 TCGCTGAGAGATATCGATCTCGGCAACACCCCATCTACCTGCTGCTGCTGCTGCTGCTGCT 2969

Oy 301 tggggcgtctgagctgcctcctagcaacgggagctcaccgcgtacgcgtcagcgtttagtgg 360
    |||||||
Db 2970 TGGGGCGTGCATGCCCTTAGCAACGGGACTCTCACCCGCTACGCGCTGACGCTTATGTTGGG 3029

Oy 361 cttagcacaactgacgccttactgttggtttgtccctcagcgtgtcggttctccgcaat 420
    |||||||
Db 3030 CTAGCCAAACTGACGCTCTACTCTGTGGTTTTCGCCCTAGGGGCTCGGTTCTCCGCAAT 3089

Oy 421 ccccgcttgcagatcgtctgtctgtctgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgc 480
    |||||||
Db 3090 CCCGCTGCGGATGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3149

Oy 481 taaggcctcaacaaatgatactacgcgctgtgaagagctgagcgttgggtgagtcgcaac 540
    |||||||
Db 3150 TACGGCCTCAACCAATGATCTACGGGCTTGAAGAGCTGGGCACTTGGGTGATCGCAAC 3209

Oy 541 tcggttcgcagctcactcctcaggttatacagctatctggtgcaaccccccaactcgtcgtc 600
    |||||||
Db 3210 TCGGTGCGGCTTCTACCTCAGGGTTTACAGCTATCTGGGCAACCCCAACCTCTCGGCT 3269

Oy 601 gctatctgtgtccgacagcagctccttcttctcagcagcagcagcagcagcagcagcagc 660
    |||||||
Db 3270 GCTTATCTGCGTGGGAGAGCTGCTTTTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3329

Oy 661 ctcccacagctgtcgtgcagatcgtcgtgcagaggtgagcagcagcagcagcagcagcagcagc 720
    |||||||
Db 3330 CTCCCAGAGCTGTGGGAGATCGCTGCGACAGGTGCGAGAGCTTATGTCTGATCTCTCACC 3389

```

```

Oy 721 tacagtcgagtgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 780
    |||||||
Db 3390 TACAGTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3449

Oy 781 ctctacttgattcaaccgcgtctacccgcacccctgagcagcagcagcagcagcagcagcagc 840
    |||||||
Db 3450 CTCTACTGCTTCAACCCCGCTCTACCCGCAACCTCGGCGAGCTGCTGCTGCTGCTGCTGCTGCTG 3509

Oy 841 ttgggtggaactagtcgcgtgctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
    |||||||
Db 3510 TTGGGTGAGACTAGTCGCGGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3569

Oy 901 cgcgtgtgtgagcactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960
    |||||||
Db 3570 CGCGTGTGTGAGCATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3629

Oy 961 tgcgtgcgagtgctgagatgatatcaagacagcgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
    |||||||
Db 3630 TGCCTGGCGGTGCTGCGAGATCATTCAGATCGGCTTGGCTGGGCTATCGGCTCCCGGCAAT 3689

Oy 1021 accgccttlaacactggttaltccctctatcaacagcgcgttlaacgcgttgaagcgc 1080
    |||||||
Db 3690 ACCGCTTTAACTGCTGTATTCCTCTATCAACAGCGCGCTTACGGCGTTGAGCGCC 3749

Oy 1081 tactcgtcccgctgagagtcgagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
    |||||||
Db 3750 TACTCCGTCCTCGTGAAGTCGCGGTTGAGGCGGACTACTGTGGCTTGAAGCGCTTGTGCT 3809

Oy 1141 tgcgtgtctgtgctgacagcagtgtaagcagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
    |||||||
Db 3810 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3869

Oy 1201 cgcacatcccaagccttltgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
    |||||||
Db 3870 CGCAATCCCAAGCCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3929

Oy 1261 caagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
    |||||||
Db 3930 CACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3989

Oy 1321 attgagcagatcgcgagcttctcgcagcccccaacacctcagaacactcctccagaagcc 1380
    |||||||
Db 3990 ATTGAGCAGATCGGAGTTTCTGCGAGCCCCCAACCTTCCAAAGCACTCCTCCAGAAAGCC 4049

Oy 1381 gagcattcagagcaaaaatcgtag 1404
    |||||||
Db 4050 GAGCATTCAGAGCAAAAATGTAG 4073

RESULT 2
D90911 LOCUS D90911 143051 bp DNA BCT 04-JUL-2001
DEFINITION Synechocystis sp. PCC 6803 DNA, complete genome, section.13/27, 1576593-1719643.
ACCESSION D90911 AB001332 BA000022
VERSION D90911.1 GI:1653083
KEYWORDS
SOURCE Synechocystis sp. PCC 6803 (strain:PCC6803) DNA.
ORGANISM Synechocystis sp. PCC 6803
REFERENCE 1 (bases 1 to 143051)
AUTHORS Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N., Sugiura,M., and Tabaata,S.
TITLE Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome
JOURNAL DNA Res. 2 (4), 153-166 (1995)
MEDLINE 96127529
REFERENCE 2 (bases 1 to 143051)
AUTHORS Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y., Miyajima,N., Hiroseawa,M., Sugiura,M., Sasamoto,S., Kimura,T., Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,

```

Okumura,S., Shimpou,S., Takeuchi,C., Wada,T., Watanabe,A.,  
Yamada,M., Yasuda,M. and Tabata,S.  
Sequence analysis of the genome of the unicellular cyanobacterium  
Synechocystis sp. strain PCC6803. II. Sequence determination of the  
entire genome and assignment of potential protein-coding regions  
DNA Res. 3 (3), 109-136 (1996)

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 143051)  
Tabata,S.  
Direct Submission  
Submitted (28-JUN-1996) Satoshi Tabata, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research, Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail: tabata@kazusa.or.jp, URL: http://www.kazusa.or.jp/cyano/  
Tel: 81-438-52-3933 (ex.2330), Fax: 81-438-52-3934)  
Potential protein coding regions were assigned on the basis of  
similarity search of the ORFs and Genemark analysis.

FEATURES  
source  
1. 143051  
/organism="Synechocystis sp. PCC 6803"  
/strain="PCC6803"  
/db\_xref="taxon:1148"  
/note="synonym: Synechocystis PCC6803"  
173. 1960  
/gene="slr2105"  
173. 1960  
/gene="slr2105"  
/note="ORF\_ID: slr2105  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAA18000.1"  
/db\_xref="GI:1653084"  
/translation="MARLTALFRYGYKAFAPATAFGAGIALGVTVGMSKLYLGLM  
AIVAFPEFLMTWLMQROSGFLGRSTOSSANVASTLAVALITLINTYAFRPVAM  
DLTENQRTLAPOSQSIYAELEOPLKWIIPDASEKDRLLQNYORSPDSFEPH  
PDRKRLVQDFNVQKGDVITEYGRKQLVNLSFQOREDLSEIQLNATAKQORE  
IPQYLVVGHGELPLNENRGIISTAVALENLVGYVQPLNVYDGGIIPEDADAVING  
PTRKFLAPEIATLKQYTDQGNLVMVSPENDAGLEDLLEPMGVSPRLVIDLSGAG  
GVGSLDPTVPIFNRYGNHPITQKLOGAIAIPLVRPVATKEQPHIQAATTLVEASDLMY  
ATOLISEDLPQNPETDLGPFDVAVALRREKEVSAKEMETGNTPPEELNPSRSP  
VSPSPSGNCTNQPENAMPENDSNQAGRESPPAGENEDSNDQAOTAOQNGDELLP  
AKAVVGSARATNAMENAVANGDIFLNSVQMLAENDQPLNRAKDPDTRRINLVGR  
RATVVGWLTWIVLPLIGLAVFTWWROR"  
/complement(2133. 2651)  
/gene="apt"  
/complement(2133. 2651)  
/gene="apt"  
/note="ORF\_ID: sl11430"  
/codon\_start=1  
/transl\_table=11  
/product="adenine phosphoribosyltransferase"  
/protein\_id="BAA18001.1"  
/db\_xref="GI:1653085"  
/translation="MDKALIRDIPDPKPEIMRDTITLINSPEGLRYITDSIVBEC  
ESQELVPDVVGMESRGFLFGMPLAYOMNAGITVRKPGKLPAPVHREVDLEGRKS  
LEIHQDAVAPHRRLIVDDLTATGTAKATGELLTLKIGCEVLGFAFTIELAALNGROC  
LPDLPIISLVEY"  
/complement(3032. 3946)  
/gene="sl11429"  
/complement(3032. 3946)  
/gene="sl11429"  
/note="ORF\_ID: sl11429  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAA18002.1"  
/db\_xref="GI:1653086"  
/translation="MILVLPWLIGAGATATFVWGPNKIKNAWDEEDKKKEEKRRKRE  
ARTEAKAIPDAVERGYEMLVEKERITSEWNOVSKNLSERKNTKLNLEEDITIK  
YIROSKITYSVMEFVDSQSLIEDHYDTRICKRMKRENNKRRKKMKKRESSSVYIS  
HIFLYTLDLKKCEKQOKLIESGMSIDKMDVDEKVTLLHFTPMGTRGSLTQOT  
ODFGADLLIMKDRKKYVIOAKRSASVSKAIOEIIIGAKGYDADIAIVYTNNSYTN

AKELAYRNKVELMDRDLIEFIESRKI"  
/complement(3994. 4209)  
/gene="ss12789"  
/complement(3994. 4209)  
/gene="ss12789"  
/note="ORF\_ID: ss12789"  
/codon\_start=1  
/transl\_table=11  
/product="resolvase"  
/protein\_id="BAA18003.1"  
/db\_xref="GI:1653087"  
/translation="MGQFERKMLERQREGVAKAEKYEARKPTSDGVEQVYTVRH  
GRKAYAVARELGISOASVYRINLWCSA"  
/complement(4340. 5218)  
/gene="sl11428"  
/complement(4340. 5218)  
/gene="sl11428"  
/note="ORF\_ID: sl11428"  
/codon\_start=1  
/transl\_table=11  
/product="P3 protein"  
/protein\_id="BAA18004.1"  
/db\_xref="GI:1653088"  
/translation="MESNRLTTFPLDALFLINFGMGIGLITIKDNRILDEPKAIVAIG  
LIAQLVLPWVGFSLSPSLPELAVGMILAACGSGTSNVITYLKGNVALSTL  
TAISLVTITITIPLVNLANVMEGEFALQLEFVTQIAVITITIPSLGIPIYF  
LPKIAMMEKNVAKLSLFLGITITVALIVERENLADPFLQVGVTVTNLITMGIGY  
GIALIAGIGLAERKSITIEVIGNGTLAIAIATTLNAPIMAIIPAIYSVWMTISGI  
FACILAKATPRGAKST"  
5614. 6606  
/gene="slr1503"  
5614. 6606  
/gene="slr1503"  
/note="ORF\_ID: slr1503  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAA18005.1"  
/db\_xref="GI:1653089"  
/translation="MTISTDTSLIIMPAGCGTWMGKLLMGYOPANNGELEVER  
EYCAATGTFPDQDSYGTGRLOSSEKILIGFSAAYEGINOKTICATLATYAPMLT  
PQAMVGAGMASOORLQRPIDLVOLHNSTANFAPWQDPLIKGLVOICQGHAKILGS  
NFPDRKQKARDFCQNEGVSIVSLQVYSLSLSDPVSKGLKELCQDQIKLIAVSP  
TLIELGKGVQDQPPFGIIGLFRRLRPRIQPLDITQAIANERKTAOVALNPCI  
CQGTMPDPCAGKNCQGVODNGLAGQWLDREMDLNNOAAALPQTVVQNLFQSR"  
/complement(6657. 7907)  
/gene="hob"  
/complement(6657. 7907)  
/gene="hob"  
/note="ORF\_ID: sl11427  
gene: hob or htrh"  
/codon\_start=1  
/transl\_table=11  
/product="protease: hob"  
/protein\_id="BAA18006.1"  
/db\_xref="GI:1653090"  
/translation="MAIHUKASHLGYAVALLLIFGCAIGAAGGYLLSSGQNHSSPDS  
VNTSPQSLTPAPVESNYRSALPLTLPRASODDGLNFIARAQKIGPAVVRIDSEPTA  
VSGQPGMGDQPFRRFREGEMPNPDRQGTGSGFLLSSDGEVLNNAVVGASVYK  
VTLAKDSYDEGKRWGIDTMTDVAVYVEAENLPVYVIGSDRLQRPENALIAIGNPLGL  
DNTVYGITISALGRSSSEVGVPDKRYRFOTDDAATNPGSGGLNAAKEVATIGVNTAI  
RADAOGLGFAIPIQTAQNAVAENLETKGMEHPYLGTHWVTLPEMTKOLRTSGELPAG  
VPAIDTGVLLIOVSPGSPAQAQGLAPGDDIILEVVGMGVKTATVVOERHVSQIGEPAL  
AVKRGKQPMMAVAPRPFPELIGQ"  
/complement(8021. 8491)  
/gene="sl11426"  
/complement(8021. 8491)  
/gene="sl11426"  
/note="ORF\_ID: sl11426  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAA18007.1"



```

/db_xref="GI:1653091"
/translation="MMKSLRGIALSVLMATGCGLTLSAOVA PGDDIPITVOEALD
TPDVIDITPOPOLITWICKNGSKSVAVKELKGOEIVNNDOTWACEDONIPSDPG
DAFSCETNATMGILITVFLBEGSGKAEITSMNNLANNNVCTHGSNPFWE"
gene
complement(8617..10470)
/gene="pros"
complement(8617..10470)
/gene="pros"
/feature="ORF ID:s111425
gene:pros of drpa"
/codon_start=1
/translation="MNDNSVRLPLRLIMRLSOML FVLRDDPADAEIPSKLLVLAG
PTRLGSGIATLPLIMRYLEKRIIVOEENKRTAOECLPOLPSELIMKSGHMDT
YTESGIMFALDRLERELGLPHEVITAIAKIMIRSYROLPNVLYOIQKPFDEI
RRFGLIMRGREFIMKDAYSFHSEDAISLATYGAHQANIRFSRGGLDPEYDADSGA
IGSSGSEOFENVLADAGEDEIYTEDGLSANERKAVSLAPDAKPSFGYKOSAPNT
ATIAKCCOICDSDANVYKAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVYVAVYV
PHYESTNLSLTPDEKEDOKKAAKPLPLGITSPEISDYIATINKOVADKELRLVDT
AVDLNPAFGADAVDYHTGANWSTDFLPKLVVDVRAKALADRPVHDPTEGLKARG
LEVGHIFQLGKYSQAMGATFTNEGEEKPLVMGCGYIGVSRSLAQAAYEOSYDKGII
MPTALAPYEVIVPNVGDDEQMOVAEOIYEFQAEVVDLDDNERAGYKFNDEL
IGIPRYVYTKSLKQGLEVYKRAKSVONLIAVAVPPLTAMIAEKKTS"
10639..11235
/gene="s1r1505"
10639..11235
/gene="s1r1505"
/feature="ORF ID:s1r1505
unknown protein"
/codon_start=1

```

```

Query Match      23.2%  Score 326.4, DB 1; Length 143051;
Best Local Similarity 52.7%: Pred. No. 6, 2e-66;
Matches 708; Conservative 0; Mismatches 636; Indels 0; Gaps 0;

QY 1 atgactgtctggcaaacctgactgcttctgacatcccaacacagtgagccacac 60
DB 28283 ATCTCATCTGCGGATCGCTGATGTTTGGCGGTTTTTCCCGCAGGAATGGGCGGCG 28342
QY 61 agttctctgacatcgctgttctgacagcctcgagcttggcgccctccacagctgt 120
DB 28343 AGTGTGCTCATCTGTTGGTGGGCTGGGACAGAGTTGATACAGGCTATGTGCTCTG 28402
QY 121 gtttggctgagagcactggtgtgtctctgtctgtctgtctgtctgtctgtct 180
DB 28403 CCCCACTTCGAGGCAATTGGGTACGGCTCTAGTGGCAATATTATTGCGGCTCCCT 28462
QY 181 gtgcacagtcgcgcctgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240
DB 28463 ACCTCCACACCATGTTGGGCAATTTTATGCTGCTGTGTGGAGCCCTTTGGGCTGTG 28522
QY 241 tgcgtgacagatagatcgtgcgaacgaaccccatcactgtgtgtgtgtgtgtgt 300
DB 28523 ACCTTGTGATCAACACAGGAGGTTTGACTCCCATCCATGTTTATGTTTTCGCTAC 28582
QY 301 tggggcgctcgatcgctgacagcgagctctaccccgtaacgcgtgtgacagctt 360
DB 28583 TGGTCAATTTTGGGCGATGCGCTGGGATTTTCTCGGTAAATGGCGGCGGCTGCGGG 28642
QY 361 ctacgcaaacactgacgtctacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
DB 28643 TTAGGGAATTAACAGCTAATTTATGTCTTTCTACTGCGCGGCAAGGTTATTCGAA 28702
QY 421 cccgctcgcgacgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
DB 28703 AACAATGTTTAAACCGGTTAATAACCGTTGTTTACTGTAGGCGCTATTGTTGGGAG 28762
QY 481 taggcctcaaacactgacgtctacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 540
DB 28763 TACGGTGTCCGACACAGGTGACGGGGGTAGACAGATTAGCCACTTTGGAATGACCC 28822

```

```

QY 541 tgggtgtgcgacttcaaccctcaacgggtttacagctatctgggcaaccccaactgtgt 600
DB 28823 TCTACCTTGCCCGACGCACTAGGGTATATATACCTTTTATAGTAAATCCAAATCTTGG 28882
QY 601 gcttactgtgtgcgacagctgcttcttctgcagcagcagatgggtgtgtgtgtgt 660
DB 28883 GCTTACCTGTGTGCCATGACGCGTTTGAGCTTGATGATGCCCTGTGTGTGTGTGTGT 28942
QY 661 ctcccaacgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
DB 28943 TGGCCCAAACTGCTGGAGACCAACCATGATGATTTAAACCTACTGCTCTTTTAA 29002
QY 721 taaagtcgagtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
DB 29003 CAGAGCCGGGGGGGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29062
QY 781 cctactgtgttcaaccccgctacccgacccctggagcagcgtgtgtgtgtgtgt 840
DB 29063 TACTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29122
QY 841 ttgggttgacatgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900
DB 29123 ATCCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29182
QY 901 cgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960
DB 29183 AGGGCCATGATGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29242
QY 961 tggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1020
DB 29243 TGGGAAGGGGTAAAGCCATGATCCGACCCGCTTATCATTTGGATTTGGATTTGG 29302
QY 1021 accgcttcaacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1080
DB 29303 GAAGCCTTTAAACCAATTTATCTTACTATATATGCGGCCCGCTTACCGCCCTG 29362
QY 1081 tactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1140
DB 29363 TATTCATTTTACCTTAAAGAAATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29422
QY 1141 tggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200
DB 29423 TGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29482
QY 1201 cgcaccccaacgcttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1260
DB 29483 CTGCGCCCGGAAGGATCTGATTTATGGGGCTTTAGCGGCGATCAGTCGTTGTG 29542
QY 1261 caggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
DB 29543 CACGGCATGTGATACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 29602
QY 1321 attgagcgtatcgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380
DB 29603 GTGGCATTTGTTGCTAGTCACTGG 29626

```

```

RESULT 3
LOCUS 166494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers

```

source 1. .7218  
/organism="unknown"  
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others  
ORIGIN

Query Match 3.5%; Score 49.8; DB 6; Length 7218;  
Best Local Similarity 8.6%; Pred. No. 0.21;  
Matches 39; Conservative 217; Mismatches 199; Indels 0; Gaps 0;

OY 86 gctcgcagcttgccgagccctccagccagctgcttgctcgcagcgaactgggtgct 145  
Db 1041 GCTCTCAGCTCAGCGAGCTTCGCATYYYYYYYYYYYYYYYYYYYYYYYYYY 1100  
OY 146 tctctctcgtctcgtctcagcttcgctcgccttcgctccagctccgacctagggttgg 205  
Db 1101 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1160  
OY 206 ggtcagcgcgatacgcgagctcatttgagccctcgtctcgtcgcagatacgcagcgcg 265  
Db 1161 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1220  
OY 266 aagcaacccccctcactgctgctgctcactgagggcgctgagccctcagaacg 325  
Db 1221 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1280  
OY 326 gactcaccgcgtaacgcgctcagctttagtctgagctagcaaacgactcactcgt 385  
Db 1281 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1340  
OY 386 tggcttctgacctagcggctcgcgaatcccgctcgcagctcgtctgtctc 445  
Db 1341 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1400  
OY 446 cgtctcgtctgctcagcttcgtctcagctgctcagctcagcaacgactcagc 505  
Db 1401 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1460  
OY 506 gctctgaagagctgcgacttggtgagtcagcgaac 540  
Db 1461 TACTTGCAATAGATAGTATATACAGTCAAGCCTAC 1495

RESULT 4  
AF106077 810 bp DNA BCT 20-APR-1999  
LOCUS Mycobacterium tuberculosis mutant NADH-dependent 2-trans enoyl-acyl  
DEFINITION carrier protein reductase (inhA) gene, complete cds.  
ACCESSION AF106077  
VERSION AF106077.1 GI:4588369  
KEYWORDS  
SOURCE  
ORGANISM  
TITLE  
REFERENCE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .810  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
1. .810  
/gene="inhA"  
/note="mutant, 1194T"

CDS

1. .810  
/gene="inhA"  
/note="isoniazid and ethionamide target protein; inhA"  
/codon\_start=1  
/transl\_table=11  
/product="mutant NADH-dependent 2-trans enoyl-acyl carrier  
protein reductase"  
/protein\_id="A026114.1"  
/db\_xref="GI:4588370"  
/translation="MTGLDGRKRIIVSGIITDPSIAFHIAVRAOEGAOVLVTFGRDL  
RLIQRITDRLPAKAPLLEDYDQNEHLASLAQRVMAICAKKLDGVHSIGFMPQTG  
MGINPEFADYADVSGKHTHSATSTASMAKALPILNPGSGIVGMDFSRMPRYNM  
MIVAKSALDSVNFVAREAGKYGVRNLIYAAGPTRLIASALVGALEAEAOIQLL  
EEGWDRAPRIGMMKMDATPVAKTVCALISDMVLPATYTDIIVADGGAHTQLL"

BASE COUNT 143 a 250 c 281 g 136 t  
ORIGIN

Query Match 3.4%; Score 47.4; DB 1; Length 810;  
Best Local Similarity 54.2%; Pred. No. 1.1;  
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

OY 1070 cgttgagcgctactcgtcccgcttggaagtcgaggttgagggcgagcttggtga 1129  
Db 56 CgtcGATCGCTTCACATCGACGGGTAGCCAGAGAGAGAGAGAGAGAGAGAGAG 115  
OY 1130 cgcgcttcgcttgctcgtctgagctcagcgcggtgagcgagtgagccgac 1189  
Db 116 CCGGGTTGCGACCGCGCTGCGGGCTGATTCAGCGCATCACCAGCGGCTCCGGAAGGCC 175  
OY 1190 tgcgagcgatcgcaatcccaacgcttgggtgagtgctgagcttgccggtttg 1246  
Db 176 CGCTGCTCGAAGCTGACGTGCAAAACAGAGACACCTTGCGCGCGGG 232

RESULT 5  
MTU02492 832 bp DNA BCT 26-JAN-1994  
LOCUS Mycobacterium tuberculosis H37Rv isoniazid and ethionamide target  
DEFINITION protein (inhA) gene, complete cds.  
ACCESSION U02492  
VERSION U02492.1 GI:407313  
KEYWORDS  
SOURCE  
ORGANISM  
TITLE  
REFERENCE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .832  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="pYUB 315"  
11. .16  
22. .831  
/gene="inhA"  
22. .831  
/gene="inhA"

1 (bases 1 to 832)  
Banerjee, A., Dubnau, E., Quemard, A., Balastramanian, V., Um, K. S.,  
Walson, T., Collins, D., de Lisle, G., and Jacobs, W. R. Jr.  
inhA, a gene encoding a target for isoniazid and ethionamide in  
Mycobacterium tuberculosis  
Science 263 (5144), 227-230 (1994)  
94112548  
2 (bases 1 to 832)  
Banerjee, A.  
Direct Submission  
Submitted (11-OCT-1993) Aseesh Banerjee, Microbiology and  
Immunology, Albert Einstein College of Medicine, 1300 Morris Park  
Avenue, Bronx, NY 10461, U.S.A  
Location/Qualifiers  
1. .832  
/organism="Mycobacterium tuberculosis"  
/strain="H37Rv"  
/db\_xref="taxon:1773"  
/clone="pYUB 315"  
11. .16  
22. .831  
/gene="inhA"  
22. .831  
/gene="inhA"



Query Match	3.4%	Score 47.4	DB 6	Length 2232
Best Local Similarity	54.2%	Pred. No. 0.9	Mismatches 81	Indels 0
Matches 96	Conservative 0			
QY 1070	cgcttgaagcgactactccgtccgcgcttgaagtcgcgagttgaagcgagactactgagcttga	1129		
DB 1311	CGTCATCGCGTTTCATCATCGACGCGGTAGCCAGAGAGAGAGGCGCCGACAGTGTGCTCTCA	1370		
QY 1130	cgagctctgcttgctgtcgtctgtcgtgaagcgagcggtgaagcgagtgcgagctgaagcgagc	1189		
DB 1371	CCGGGTTCACACGCGGCGTGGCGGTGATTCACGCGATCACCGACGCGGCTGCCGCAAAAGGCC	1430		
QY 1190	tcgcgcgcgcatcgcaatccccaagcctttgttgtagtgctgagcttgagcggttttg	1246		
DB 1431	CGCTGCTCGAACTCGACGTGCANAAACGAGAGACACCTGGSCAGCTTGGCGCGCGGG	1487		
RESULT 8				
LOCUS AX111368				
DEFINITION Sequence 2101 from Patent WO0123604.				
ACCESSION AX111368				
VERSION AX111368.1				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
source				
BASE COUNT				
ORIGIN				
Query Match	3.4%	Score 47.4	DB 6	Length 2609
Best Local Similarity	54.2%	Pred. No. 0.87	Mismatches 81	Indels 0
Matches 96	Conservative 0			
QY 1070	cgcttgaagcgactactccgtccgcgcttgaagtcgcgagttgaagcgagactactgagcttga	1129		
DB 1855	CGTCATCGCGTTTCATCATCGACGCGGTAGCCAGAGAGAGAGGCGCCGACAGTGTGCTCTCA	1914		
QY 1130	cgagctctgcttgctgtcgtctgtcgtgaagcgagcggtgaagcgagtgcgagctgaagcgagc	1189		
DB 1915	CCGGGTTCACACGCGGCGTGGCGGTGATTCAGCGCATACCGACCGCGTGGCGCAAAAGGCC	1974		
QY 1190	tcgcgcgcgcatcgcaatccccaagcctttgttgtagtgctgagcttgagcggttttg	1246		
DB 1975	CGCTGCTCGAACTCGACGTGCANAAACGAGAGACACCTGGSCAGCTTGGCGCGCGGG	2031		
RESULT 9				
LOCUS I26125				
DEFINITION Sequence 1 from patent US 556778.				
ACCESSION I26125				
VERSION I26125.1				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
1 (bases 1 to 3120)				

AUTHORS	Sacchettini J.
TITLE	Crystalline Inha Enzyme-NADH Complex
JOURNAL	Patent: US 5556778-A 1 17-SEP-1996;
FEATURES	Location/Qualifiers 1..3120
source	/organism="unknown"
BASE COUNT	566 a 995 c 1037 g 522 t
ORIGIN	
Query Match	3.4%; Score 47.4; DB 6; Length 3120;
Best Local Similarity	54.2%; Pred.No.0.85;
Matches	96: Conservative 0; Mismatches 81; Indels 0; Gaps 0;
OY	1070 cgcttagcgccactaccgctccgcgtgtgaagtcggcttgaaggcggactacttggtcta 1129
Db	1041 CGTCGATCGCGCTTWCACATCCGCCGGGTAGCCACGAGCAGAGGCGCCGCACGTGTCATCA 1100
OY	1130 cgagcttgtctgtgtctgtctgtgtcaaggcggtgaacggtggtggtgagatgtgacgac 1189
Db	1101 CCGGGTTGACACCGCGCTGGCGGTGAATTCAACGCCCATCACCGCGCTGCCCGCAAAGGCC 1160
OY	1190 tcggcgcgcatcgcaatccccaaagcctttgttatgtgatgtactgtgcggttttg 1246
Db	1161 CGCTGCTGAACTCGACGTGCATAAACGAGAGACACTGGCCAGCTTGCGCGCGGG 1217
RESULT 10	
LOCUS	I73545 3120 bp DNA PAT 03-APR-1998
DEFINITION	Sequence 11 from patent US 5686590.
ACCESSION	I73545
VERSION	I73545.1 GI:3009686
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified. 1 (bases 1 to 3120)
AUTHORS	Jacobs,W.R., Jr., Collins,D.Michael, Banerjee,A., de Lisie,G.William and Wilson,T.Mary.
TITLE	Methods and compositions for detecting and treating mycobacterial infections using an INHA gene
JOURNAL	Patent: US 5686590-A 11 11-NOV-1997; Location/Qualifiers 1..3120
FEATURES	source /organism="unknown"
BASE COUNT	566 a 995 c 1037 g 522 t
ORIGIN	
Query Match	3.4%; Score 47.4; DB 6; Length 3120;
Best Local Similarity	54.2%; Pred.No.0.85;
Matches	96: Conservative 0; Mismatches 81; Indels 0; Gaps 0;
OY	1070 cgcttagcgccactaccgctccgcgtgtgaagtcggcttgaaggcggactacttggtcta 1129
Db	1041 CGTCGATCGCGCTTWCACATCCGCCGGGTAGCCACGAGCAGAGGCGCCGCACGTGTCATCA 1100
OY	1130 cgagcttgtctgtgtctgtctgtgtcaaggcggtgaacggtggtggtgagatgtgacgac 1189
Db	1101 CCGGGTTGACACCGCGCTGGCGGTGAATTCAACGCCCATCACCGCGCTGCCCGCAAAGGCC 1160
OY	1190 tcggcgcgcatcgcaatccccaaagcctttgttatgtgatgtactgtgcggttttg 1246
Db	1161 CGCTGCTGAACTCGACGTGCATAAACGAGAGACACTGGCCAGCTTGCGCGCGGG 1217
RESULT 11	
LOCUS	I86995 3120 bp DNA PAT 10-JUN-1998
DEFINITION	Sequence 1 from patent US 5702935.
ACCESSION	I86995
VERSION	I86995.1 GI:3206713

KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN
unknown.	unknown.	unknown.	Unclassified.	1 (bases 1 to 3120)	Sacchetti, J., Blanchard, J., and Jacobs, W.R. Jr.	Method and compounds for inhibiting lipid biosynthesis of bacteria	Patent: US 5702935-A 1 30-DEC-1997;		
						Location/Qualifiers	1..3120		
						/organism="unknown"			
								566 a	995 c 1037 g 522 t
Query Match			3.4%;	Score 47.4;	DB 6;	Length 3120;			
Best Local Similarity			54.2%;	Pred. No. 0.85;					
Matches 96;			Conservative 0;	Mismatches 81;	Indels 0;	Gaps 0;			
Qy 1070	cglttagcgccctacccgcgcgcgtcggaagtgcggtltagagcgcgagcttgcgttga	1129							
Db 1041	cgttcattgcgctttacatcattgcacgacggatgcccagacagacggccgccacgttggtctta	1100							
Qy 1130	cggccctcgcttgcgtctgtctgtgtgtacagcgcggtltagcgcgagcttgacgcagc	1189							
Db 1101	ccgggttcgacgcgcgtcggtgatgttcagaccatccagacgcagaccggcggcggcggcc	1160							
Qy 1190	tgcgagcgagtcgcacatcccagaagccttgglttatgctatgcttgccgggttgg	1246							
Db 1161	cgctctctgaactcgacgtgcnaaaacgagagacacccactgcccagcttggccggcggg	1217							

/gene="MT1526"  
 /note="similar to GP:5458705; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="magnesium chelataase, putative"  
 /protein\_id="AAK45791.1"  
 /db\_xref="GI:13881146"  
 /translation="MTSAGCFPACAGGYQTPEGSHASAPANEAPPGEAGLAEVHTLE  
 RAIFEKRIALQDOLVSRMIVGLTSKCVLLECGVAKTIAVETFARVVGGSFST  
 OETPLDIPVDITIGRTVIRGOREEPTDELGPVAVNADLEINRAPAKVGSLLEWME  
 RVSTVIGGRFPEPSPETLYMATQNPTEHGCVYIPEAORDREITNGCYSPSEERI  
 IYRMGYTPPOAKOILSTDGLRLDEIANNVHHALVDYVYRVYFATRRKEDQANDV  
 KSWAVGASPRASLGTIAAASLALVRSDVTVODYIEVLPDLRLRLVITYDALD  
 EISPELVINLQTVLALPOVNAVNPDOOGSVPPVMAAAAAASGR"  
 1272..2183  
 /gene="MT1527"  
 1272..2183  
 /gene="MT1527"  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAK45792.1"  
 /db\_xref="GI:13881147"  
 /translation="LTRGGDIDDPKIAALRTLELTIVKQKLDGVLRHGHLITPGPGE  
 PEESRIYQPGDVRMDMAVYARTHTHPYRQMIADRELETRLVYDKMSASLDFTACCE  
 KADLVAAALATFLNSGGNRNLGALIANGAAMRPVAPRTGRKQTHMLRTIATMPQA  
 PGVGLVDVLAIDALRPERRRGMAVVISDFLGIINMRPLRAIAAHNEVLAIEVDP  
 RQVELPDMDVADLVQAESGVVRESIDPLDFAAAAAAHRAVDARTINGCGAPILS  
 LRTDDPDVADLVIEFVASRRRCAIAGHQ"  
 1272..2183

RESULT	12
LOCUS	AE007022
DEFINITION	Mycobacterium tuberculosis CDC1551, section 108 of 280 of the complete genome.
ACCESSION	AE007022
VERSION	AE000516
KEYWORDS	AE007022.1 GI:13881145
SOURCE	
ORGANISM	Mycobacterium tuberculosis CDC1551. Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE	1 (bases 1 to 20113) Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Salzberg, S.L., Delcher, A., Umayam, L.A., Ermolaeva, M., Gill, J., Mikula, A. and Bishai, W. Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
TITLE	Unpublished 2 (bases 1 to 20113) Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W. Direct Submission Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
JOURNAL	
REFERENCE	
AUTHORS	
FEATURES	
source	Location/Qualifiers 1..20113 /organism="Mycobacterium tuberculosis CDC1551" /strain="CDC1551" /db_xref="taxon:83331" /note="clinical strain" 100..1233 /gene="MT1526" 100..1233 CDS

CDs

gene

complement(3274. .4152)

/gene="MT1529"

complement(3274. .4152)

/gene="MT1529"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAK45794.1"

/db\_xref="GI:13881149"

/translation="MTPELVGQHTSTMDPRLSGEALACGLTPYEELRSRYVALHNDY  
IVPGVEELTAEQAKAMLRSSRRRGVLAGYSASAFHGAKIMDADLPAAILIDITRRRAP  
GLQWEEIIPDEICVIEGMRTPTPERIALDLSPFLDPVAVAUVALIOATLIKVD  
VEPLLEERGRGRCKAARAALDLVDGSQSKEETMLLLIRAGFRPOTOLVRNEM  
GMAEHLMGMODIKVAAREYDGDNHLISRYHYRKDILRHKKYQNRGWMIVVVAYEDH  
PADILRRVGEARAFRA"

4257. .5000

/gene="MT1530"

4257. .5000

/gene="MT1530"

/note="similar to PID:1155269; identified by sequence  
similarity; putative"

/codon\_start=1

/transl\_table=11

/product="3-oxoacyl-(acyl-carrier-protein) reductase"

gene

CDs

gene

complement(3274. .4152)

/gene="MT1529"

complement(3274. .4152)

/gene="MT1529"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAK45793.1"

/db\_xref="GI:13881148"

/translation="MTPELVSPMILTSFGHSWFPLFFVYAGLYALYLMOQLQRMM  
LRFAMEELLEESAPKRPSRMRHPVALILVLSLLFTILAMGPPTHDAVIPRNRAVMVLY  
IDVOSMSMATIDEPSRMVAQAQEAARQFADDELTPGINGLIATPATAVILSPTNNREA  
TNALDKIGTFADRATGTGEAIFLTALQAIATVAVAGGGDIPTXPARIVLDSQKETPMTN  
PNNGPQTAAARTAKDGCVPISTISFGPQVEEINOROPVPVDDTEFMKKVWLDSGG  
NSNYNAITLAELRAYSSLDQDIGETTKGDASVGMILGALATALMALDALILNRRLP  
"

gene  
CDS  
/protein\_id="AAK45795.1"  
/db\_xref="GI:13881150"  
/translation="MTATRTSGAMPPEYRSVLYTGCNRGIGLAIQRLAADGKAVV  
THRRSGAGLEGVECDVTDSDADVRATVAEEHQGPVEVLVSAGLSADAFMKRME  
EKFEYFVNANLITGAFRVAQRASNRQNRKFGMRITIGSGSGSWGIGMDANVAAKSGAV  
IGMASIRRELSKANVTANVAVPGVITDDMTALDERIQGALGFIPAKRGTGPAEYVA  
GVSVFLASDDASYSIGAVIPYDGKMGMH"  
5019..5828  
/gene="MT1531"  
5019..5828  
/gene="MT1531"  
/note="similar to GB:U02530 SP:P42829 PID:437204;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="enoyl-(acyl-carrier-protein) reductase"  
/protein\_id="AAK45796.1"  
/db\_xref="GI:13881151"  
/translation="MTGLDGRKRLVSGIITDSSIAFHIAKVAOEOGALVTGFEDRL  
RLIOGTFRPLPAKAPLLELDVONEEHLASLQAVTEALGAGNKLQGVVHSIGFMPQOYV  
MCINFFPAPADVADKGIHTISAYSASNAKALPLPMNGSGITGMDPDSAMPAYNM  
MVAASALESVNRFPARERAGKYGSNSLVACGPIRTAKMSIYVGCALGEPAGAOIQLL  
EGMDQRAPLGKMMKDATFPKATVCALLSDMLPATGTGDIITADGAHTQL"  
5834..6868  
/gene="MT1532"  
5834..6868  
/gene="MT1532"  
/note="similar to P1D:2347006; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ferriochelatase"  
/protein\_id="AAK45797.1"  
/db\_xref="GI:13881152"  
/translation="MQQPAVLLISGSGPSPQVRYPLENTYRGRGVPAERLDVAHPE  
YHRCGVSPIKNTLTIAELAQOELPVYRGNRMPEYVDATAMRONCVRRAAF  
ATSAKSGYSCCTQYVEDIARARRAGDAPLVELKRLPTFDLPLEVEHPADAITAAFT  
VQGDRLVFTHASIPTAADRCRGPMLYSRQVAVATRLVAAAGCDEPLDQMSGGP  
QVPMLEPPTDQGLGAGINAVIVCIGVADHIEVMQDLDELRLQADAQIAVAV  
RSTPNADPRLGRLARGLIDELRYGRIPARISGDPVPVCGCLSSINGCPRCRPHCVASV  
SPARPSGSP"  
gene  
CDS  
complement(6834..7700)  
/gene="MT1533"  
complement(6834..7700)  
/gene="MT1533"  
/note="identified by glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK45798.1"  
/db\_xref="GI:13881153"  
/translation="MWCPSVLSISIMANWLAGKAAPDDVDLALSIMAPTSOAVAYDAV  
DALHGELMPDVHDAGIVSLQTLAAVGRRLRKTINVLVPEGDVKGALSWAGQFEPH  
PAGHGVFIVANPEDPGSAVGLVPEFSGVDVDEAASQEPITPEICALSNVYSLPAPAF  
PVLHEYLGDDEAYLRSVARSAAEALSTIGSSSVNAPRGLVEQLLESSRQHPDPH  
APSRALYVENNAHVDALIVASAGLSRLPIGTQSLDQNRQRTALRLPLTAVVRSARMS  
AVTALHSAWPD"  
gene  
CDS  
7758..8192  
/gene="MT1533.1"  
7758..8192  
/gene="MT1533.1"  
/note="identified by glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK45799.1"  
/db\_xref="GI:13881154"  
/translation="MPVALIVLIALVVGAEALTDGWEFLMLGALAAVSYSSLLAA  
WMWMDAGVAFELIVSLILVIVPAVARRRLTQTKGIGIALGKKANVLGRVARDDG  
QYKIDGQWTAFLRPLNDGVEFERGDSVTYVQIDGATAYFKDY"  
8214..9359  
/gene="MT1533.2"  
8214..9359

	gene	/gene="MT1533_2"	/note="identified by match to protein family HMM"
		/codon_start=1	/transl_table=1
		/product="SPFH domain/Band 7 family protein"	/protein_id="AAK45800.1"
		/db_xref="GI:13881155"	/translation="MCGAGVGLFALVLPAITIVAKSVALLPOAFAVIENLGRYS RTVSQGLTLPLVPEDIRVARVDLERVSPPOPVLTEDMLTINIDRVEYDTVGQA AVETISNITGVDEDTTTLLRNRYGGFTLEOTLSRQINAQLRGVIDEATGMRGLRV ARVELRLSDPPSLIQASMERKOMADRKRKMILITAESTRIKAIKQAGCQQAOILAE GAKCAAILAAEDADROSMLRAGOSERRAAYIQAOGQAAIKDKTPAIIKAGPTSEMAY OYLDTLEDMAKGDAKNWVVPSDNNALOCTFLTRIGPDGDGYREPSPEVEDOPKHA ADEGDADVAGWFSTDDPDSIARAVAATAEALARKRPVESLSGTPRLTQG"
		9369..9725	
		/gene="MT1534"	
Query Match	Best Local Similarity	3.4%; Score 47.4; DB 1; Length 20113;	
		Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;	
Qy	1070	cgttagagcgcctaccgctccgctcgacggagatcgsggttgaaggcgactattggtttga 1129	
Db	5074	GCTGCATACGCCGTTTCACATCCGACGGGTAGCCCAGACAGACCAGCCCCCATGGTGCTCA 5133	
Qy	1130	cgagcttcgcttgctgctgtcgttcgaacggcggttacggcggttcgcgcagaagttagacgcgac 1189	
Db	5134	CCGGGTTTGACCCGGCTGGCGGGTGATTGACGGCATCACGCAGCGGCTGCCCGCAAAGCCCC 5193	
Qy	1190	tgcggcgcatccgcaalcgccaaagccttgttgatgtcgtacctgtggcggtttgg 1246	
Db	5194	CCTCTCTCGAACCTCGACGACGAAGAAGGAGGACACCTTGCGCGCGGCGGG 5250	
RESULT 13			
LOCUS	MTCYC277	38300 bp DNA BCF 03-AUG-2001	
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 65/162.		
ACCESSION	Z79701 Akl23456		
VERSION	Z79701.1 GI:3261635		
KEYWORDS	.		
SOURCE	Mycobacterium tuberculosis H37Rv.		
ORGANISM	Mycobacterium tuberculosis H37Rv. Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.		
REFERENCE AUTHORS	1 (bases 1 to 38300) Cole,S.T., Brosch,R., Parkhill,J., Garler,T., Church,C., Harris,D., Gordon,S.V., Eigleier,K., Gas,S., Barry III,C.E., Tekela,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsbey,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.P., Rogers,J., Rutter,S., Seeger,K., Skellton,S., Squires,S., Sqaes,R., Sulston,D.E., Taylor,K., Whitehead,S. and Barrell,B.G.		
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence		
JOURNAL MEDLINE REFERENCE	Nature. 393 (6685), 537-544 (1998)		
AUTHORS TITLE	2 (bases 1 to 38300) Parkhill,J.		
JOURNAL	Direct Submission Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moléculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75224 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On Jun 27, 1998 this sequence version replaced gi:1524225.		
COMMENT	Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have		

been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TB genes implemented in TlpParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

## FEATURES

```
source
    location/Qualifiers
    1..38300
    /organism="Mycobacterium tuberculosis H37Rv"
    /strain="H37Rv"
    /db_xref="taxon:83332"
    <1..37608
    /organism="Mycobacterium tuberculosis H37Rv"
    /strain="H37Rv"
    /db_xref="taxon:83332"
    /clone="Y277"
    /complement(55..1074)
    /gene="Rv1482c"
    /complement(55..1074)
    /note="Rv1482c"
    /note="Rv1482c" (MTCY277.03c), len: 339. Highly similar
    toM. avium protein TR:007396 (EMBL:AF002133) MAV346. (346
    aa), fasta scores: E(): 0, 65.2% identity in 342 aa
    overlap, slight similarity to GRPE_ECOLI P09372 heat shock
    protein grep (heat shock P (197 aa), fasta scores, opt:
    139, E(): 0.012, (28.3% identity in 159 aa overlap). Also
    similar to Rv3517(MT023.24), 59.0% identity in 273 aa
    overlap.
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv1482c"
    /protein_id="CAB02032.1"
    /db_xref="GI:1524228"
    /db_xref="SPTREMBL:P71763"
    /translation="WGVPFLQWLMQSPQPIVSPRRNRYVHENVAAISGNRTG
    VTLVFEVQHTSTMDPFLGSELAAGVLPYELRSRYVALHADVPPGVETLAQ
    LRAKALMRSRRRGLAGYSASFAGAKMIDADLPALITDNRRAPELOWEERIEP
    DEICVTEGMRTPTPTALDLSREPLDPAVAADLPATDPAKADVEPLIERYGR
    RGMKARAALDIYDGAQSPKETWLRLLIRAGFRPPTQIAVRNEMGAELHDMGM
    ODIKVALEYDGDHILTSRYHRYKRDILREHEKVOHRYGWLVRVAEDHPADITIRYGEA
    RAIFRA"
    RBS
    1024..1029
    /note="putative RBS, GATAGG, for Rv1483"
    /gene="Rv1483"
    1038..1781
    /gene="fabG1"
    1038..1781
    /gene="fabG1"
    /note="Rv1483, (MTCY277.04), fabG1, len: 247. fabG1,
    probable 3-oxoacyl-[acyl-carrier protein] reductase,
    identical to M. bovis G1155269 putative ketoacyl acp
    reductase (247 aa), strong similarity to FABG_ECOLI P25716
    3-oxoacyl-[acyl-carrier protein] reductase (244 aa), fasta
    scores, opt: 664, E(): 6.8e-35, (44.4% identity in 241 aa
    overlap; contains PS00061short-chainalcohol dehydrogenase
    family signature"
    /codon_start=1
    /transl_table=11
    /product="fabG1"
    /protein_id="CAB02033.1"
    /db_xref="GI:1524229"
    /db_xref="SWISS-PROT:O48930"
    /translation="WTATATGAKPPVSRGVLVTCGNRGITGLIAAQLADGKVAAY
    THRGSGAKGLFGVECDYDSDADRAFTAEHOGPREVYVSNAGISADFLMKHTE
    EKREKVTANLTGAFRAVQARSRSKQNRKFRMTFTISGSGMIGNANYSASRAGY
    IGMASITARELSKANVTANVAVPGYITDDMRALDERIQGALQIFPAKRVGPFAEVA
    GVSVFLASDASYISGAVIPYDGGMGKH"
    misc_feature
    1455..1541
```

```
gene
    /gene="fabG1"
    /note="PS00061 Short-chain alcohol dehydrogenase family
    signature"
    1790..1794
    /note="potential RBS, AAGGA, for Rv1484"
    1800..2609
    /gene="inhA"
    1800..2609
    /gene="inhA"
    /gene="inhA"
    /note="Rv1484, (MTCY277.05), inhA, len: 269. identical
    toINH_A_MYCTU P46533 enoyl-[acyl-carrier-protein] reductase
    andto M. bovis G1155270 enoyl acp reductase"
    /codon_start=1
    /transl_table=11
    /product="inhA"
    /protein_id="CAB02034.1"
    /db_xref="GI:1524230"
    /db_xref="SWISS-PROT:P46533"
    /translation="MTGLDCKRTIVSGITDSSIAFHIAVVAQEGCAOLVTCFDR
    RLTIORITDRLPAKAPLLELDYQNEEHSLASLARTETAIGAKNRIDGVYHSGFRPQ
    NGINFPDAPYADYSGKIHISAYSASAKALPRTINFGSIVGMDPSPKAPATNM
    MTFVAKSALESYVNRVAREKRYGVRSNLVAAGPIRTLAASAIVGALGEGEAGQIQL
    EEGMDORAPIGMNMKDATPVAKTVCALLSDWLPATGDIYADGAHTOLL"
    2615..3649
    /gene="hemZ"
    2615..3649
    /gene="hemZ"
    /note="Rv1485, (MTCY277.06), hemZ, len: 344.
    ferredoxinase similar to e.g. HEMZ_BACSU P32396
    ferredoxinase (310 aa), FASTA scores, opt:490, E():
    2e-24, (30.2% identity in 295 aa overlap)"
    /codon_start=1
    /transl_table=11
    /product="hemZ"
    /protein_id="CAB02035.1"
    /db_xref="GI:1524231"
    /db_xref="SWISS-PROT:P71765"
    /translation="MORDAVILLSPGCEPDEOVRPPLFENTRGVPAERLDAVNEH
    YLHRGVSPTINGIRITLAELEAQQELPVTFGNMNEPVVEDATVARDKGVRAAYF
    ATSAWSGSSCTQYVEDIARARAAGADAPLVKRLRYFDHPLFVEMFAADIAAAT
    VGRGARLVFTNHSIPTAADRCGNLYSRQVAYTRLVAAAGYCDPDLAQMSRSGP
    OVPALEPDVTQSLTGAGAGINAVTCCPGVADHIVVMDLHELRLOAEAGIAYA
    RASPNMDPRARLARGLIDELRGRIPARVSGPDVPQGLISINQPCRPPOHVASV
    SPAPPSGSP"
    complement(3615..4523)
    /gene="Rv1486c"
    complement(3615..4523)
    /gene="Rv1486c"
    /note="Rv1486c" (MTCY277.07c), len: 302. Highly similar
    toM. avium Y507_MYCAV 007402 hypothetical 33.5 kD protein
    mav321 (320 aa), fasta scores, opt: 1217 z-score: 1502.0
    E():0, 71.1% identity in 315 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein Rv1486c"
    /protein_id="CAB02036.1"
    /db_xref="GI:1524232"
    /db_xref="SWISS-PROT:P71766"
    /translation="MIRVSRYEQJHCKIMMCPYSYSIMANWLAKAABDDVALS
    LMAPTGSAVAAVDAAAGHTGLPMPDVHDAGVSLIQLTGRAVGRRLRGITINVLPP
    GDVRLAGVAGTQGFEDALAGEAVTVVAREPDGSAVGLVPEFSGVDDEAQSRELPPE
    LCALSMMVYSIPCARVLEHVEIGDAEALSAVSASAEALSTGLSSSDVAKRGGLVE
    QLLEVSQRHVRPDPHAPSRALVLENAHVATIIAASGLSRPLPTGTSISDAORATDA
    LRPLTAVRSARSAVTAIIHSAMPD"
    4539..4973
    /gene="Rv1487"
    4539..4973
    /gene="Rv1487"
    /note="Rv1487, (MTCY277.08), len: 144. Highly similar to
    M. avium TR:007404 (EMBL:AF002133) MAV145. (145 aa),
    fasta scores, opt: 667 z-score: 1148.3 E(): 0, 72.5%
    identity in142 aa overlap"
    /codon_start=1
```







```
gene
CDS
    /gene="mlr2113"
    5065..5808
    /note="hypothetical protein"
    /protein_id="BAB49323.1"
    /db_xref="GI:14022715"
    /translation="MPKRYLANLRRLVSGHSAVAVCGRLMMNPTQEPERYLOGSV
    PNKATRLICTYFRIDEDDLRAPEVPAPRHKALLPHOTLYENMRGAPAPVAGGT
    FTYETVPDRDILLVRSVTEFVRREAEIVTFEVRVMAEGHGGARVAGWHYGAISRL
    NMITYGAINRRQTRREPSVMTVQMAPESDPVLVGKAVYLTCGSPACVSYIMKROGRIS
    LROYLGCPRVSLDDPRLDQVAVSLVREG"
    5859..6530
    /gene="mlr2114"
    5859..6530
    /gene="mlr2114"
    /codon_start=1
    /transl_table=11
    /product="probable transcriptional regulator"
    /protein_id="BAB49324.1"
    /db_xref="GI:14022716"
    /translation="MUNSEDYGGKEADTVHALEDDIIFGRLAGTQLVEDVLIARPV
    SRHTIROLALYOLERLGIYTERNNKGAWVRRLSEEVQIYEVREMLQROALMIPLV
    SEALIAQLEETIHRISGHDVAGYLRGIEHNDREHILMFACGNAYLVASSIEHYMRIS
    LPVRANSLADREKLDVGRQHMMATEALKRRDMVLAQLCVDHLOPSKVPLYQETETA
    PAATG"
    complement(6597..7601)
    /gene="ml12116"
    complement(6597..7601)
    /gene="ml12116"
    /codon_start=1
    /transl_table=11
    /product="probable zinc-binding oxidoreductase"
    /protein_id="BAB49325.1"
    /db_xref="GI:14022717"
    /translation="MMKAFVVDKYQKKGALRLAEMPELRLNDVLEIHAAGVNL
    SKLDEGFKPILPRPPIILGHVAGIVRAGSKARFKLGDEYARSRDRIQTFAO
    SIAMDEADVALIKPKNLSMEASIPLVGLAMOVLYETAGLKKGOKVEIOAGSGVGT
    FAIQLAKIGATVATTAANADIVKIGADIYIDYKKDDEFKTIQGYDVVLSODAK
    TLEKSLHVLKRGKGLISISGPPDPFAGKOKLNNVLRKLVRLLSRGIRSKARAGVY
    SFLPMAOGDOLGETISRIESGFLRPVIDRIFFPEKTNEALATVEAGRAKGVVVKR"
    complement(7723..8538)
    /gene="ml12118"
    complement(7723..8538)
    /gene="ml12118"
    /codon_start=1
    /transl_table=11
    /product="oxidoreductase"
    /protein_id="BAB49326.1"
    /db_xref="GI:14022718"
    /translation="MNRNSGKTATVGTASSGIGRASAEIIRAGCTVTCRRKASG
    SNGISMLVCDVTDASVSLVSTVLGKTRIDLNVNAGLGLGAEESVSOSQALF
    DVNLGVMRMINAVLPSMRRLGEGRIIVNISILGLVPAPYSAAHSAVKHALEGYESL
    DHEVRANIRVSVLEPVRVTFDQNGFEPDSVAKVYDEARAGIEALLRDVMPKADLP
    EVAKVVLRAATDRHPLRRTYAGKAQVGMRLREAFPTGMFDKILRKQRLPY"
    8432..8617
    /gene="msr2119"
    8432..8617
    /gene="msr2119"
    /note="unknown protein"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB49327.1"
    /db_xref="GI:14022719"
    /translation="MPKTVKPARAKISALARPJPEAPVTIAVPEPRFIVKAPNVKG
    LACDRRLNYDRNITLCS"
    8646..9233
    /gene="mlr2120"
    8646..9233
    /gene="mlr2120"
    /codon_start=1
```

```

/transl_table=11
/product="probable transcriptional regulator"
/protein_id="BAB49328.1"
/db_xref="GI:14022720"
/translation="MRIDKRDASRSRIMEVASHRRFRDGIASGLASIMSDAKMTN
GAFYPHFSKADLVRESMASALETOSODLOQALASGGLAMAMTISPENRDPQTGC
ASALLPEIARQPPETRSIYTEKLLTLVHQLAKDLPPTRPDEGVALGVFALLGALQ

Query Match      3 0%; Score 42.8; DB 1; Length 346294;
Best Local Similarity 56.3%; Pred. No. 4.7;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1055 aggcgcgccttaagcggttagagccctactcgcctccgccttgaaagtcgcggttagagcg 1114
    |||  ||  ||  |||  ||  ||  ||  |||  ||  ||  |||  ||  ||  ||  ||  ||
Db 7925 AGGCGATTCGCTTCGCGACGTGATGCCAAGGCGCATCTGCCGGAAAGTGATGGCAAGGTG 7866

QY 1115 gactactgggcttgagcgcttcgcttgctgctgctgctgctgaagcggtgaaagcggtgc 1174
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 7865 TTCTGCGGCGCTGCGACCGACCGGCATCCGCTGCGGCGCTACACCGCGGCAAGCGAGCGC 7806

QY 1175 ggcaggtgagcgagcgcactgcgcgcg 1196
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 7805 GGCAGGTAAACCATGCTGCGCGG 7784
```

Search completed: January 10, 2002, 22:26:09  
Job time: 3192 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:01:57 ; Search time 1154.57 Seconds

(without alignments)  
13067.266 Million cell updates/sec

Title: us-09-887-038-2

Perfect score: 1404

Sequence: 1 atgactgtctgcgaactct.....attcagacgaaaaaatgtag 1404

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues 22703874

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: qb\_est1:\*  
11: qb\_est2:\*  
12: qb\_hic:\*  
13: qb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pin:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	52.4	3.7	925	13	CNS0091P
2	47.2	3.4	1101	13	AL1084630 Drosophila
3	46.2	3.3	500	10	BE602908 HYSMEH010
4	43.8	3.1	645	13	CNS01213
5	43.4	3.1	922	13	CNS0073W
6	42.4	3.0	604	13	AZ935264 BU_Ba000
7	42.4	3.0	925	13	CNS0091P
8	42.4	3.0	1201	13	CNS016B0
9	42.2	3.0	970	13	CNS010C9
10	42.2	3.0	997	13	CNS006DN
11	42	3.0	1250	10	BE545125
12	41.8	3.0	846	13	CNS010R1

C 13	41.8	3.0	1201	13	CNS01679	AL106383 Drosophila
C 14	41.8	3.0	1868	11	BG167536	BG167536 602342745
C 15	41.6	3.0	1100	13	CNS016KD	AL106855 Drosophila
C 16	41.2	2.9	1101	13	CNS01258	AL101954 Drosophila
C 17	41	2.9	446	11	BG274660	BG274660 WHE237_E
C 18	40.8	2.9	464	10	AI858018	AI858018 wj69c12.x
C 19	40.8	2.9	663	13	CNS03268	AL1224801 Tetradon
C 20	40.6	2.9	912	13	CNS006M3	AL065775 Drosophila
C 21	40.2	2.9	914	13	CNS00CZP	AL059740 Drosophila
C 22	40	2.8	339	10	AW582345	AW582345 MR3-ST022
C 23	40	2.8	827	11	BG845859	BG845859 1024011E0
C 24	39.8	2.8	844	13	CNS0052P	AL056552 Drosophila
C 25	39.8	2.8	977	13	CNS014H2	AL104144 Drosophila
C 26	39.8	2.8	1101	13	CNS00LXT	AL078875 Drosophila
C 27	39.6	2.8	910	13	CNS006DN	AL056569 Drosophila
C 28	39.6	2.8	937	13	CNS006XP	AL066056 Drosophila
C 29	39.2	2.8	764	11	BI151035	BI151035 602916857
C 30	39.2	2.8	1159	13	CNS015XK	AL106041 Drosophila
C 31	39	2.8	442	10	BE725159	BE725159 894081D05
C 32	39	2.8	839	13	CNS004NB	AL054280 Drosophila
C 33	38.8	2.8	550	11	BI181863	BI181863 UML-P-FN-
C 34	38.8	2.8	1101	13	CNS016HG	AL106750 Drosophila
C 35	38.6	2.7	1101	13	CNS00E0E	AL067626 Drosophila
C 36	38.6	2.7	1101	13	CNS00EAX	AL068833 Drosophila
C 37	38.4	2.7	305	10	AV631321	AV631321 AV631321
C 38	38.4	2.7	407	11	BG908379	BG908379 Talrj167F
C 39	38.2	2.7	431	10	AL562202	AL562202 AL562202
C 40	38.2	2.7	972	13	CNS02BJJ	AL189928 Tetradon
C 41	38	2.7	1101	13	CNS014Y2	AL104736 Drosophila
C 42	37.8	2.7	765	11	BF865981	BF865981 963063E12
C 43	37.6	2.7	321	11	D24188	D24188 RICR1503A.R
C 44	37.6	2.7	472	11	BG464251	BG464251 EMI 71.E0
C 45	37.6	2.7	681	10	AI403917	AI403917 GH23463.5

#### ALIGNMENTS

RESULT 1  
LOCUS CNS0091P 925 bp DNA 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL053013 GI:4934461  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Drosophila melanogaster  
fruit fly.  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> the BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's Laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp. the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_pac.htm](http://bacpac.med.buffalo.edu/drosophila_pac.htm).



Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rving@clemson.edu  
Seq primer: ATTATACCTCCTAAAGG  
High quality sequence start: 3  
High quality sequence stop: 472.  
Location/Qualifiers

## FEATURES

source

```
1..500
/organism="Hordium vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEH0100N13f"
/clone_lib="Hordium vulgare 5-45 DAP spike EST library
HYCDMA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLK"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      97 a      137 c      162 g      103 t      1 others
ORIGIN
```

Query Match 3.3%; Score 46.2; DB 10; Length 500;  
Best Local Similarity 53.3%; Pred. No. 0.35;  
Matches 96; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
QY 1057 gcgcgttacgagcttgagcgctactccgcgtgaaagtcgagtgagggcga 1116
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 GCGGCGCGCGCGCGGCAACGAGCGGACCGCCGCGCAACGAGGGGTGAGAGAGG 131

QY 1117 ctactgagcttacgagcttgctgctgctgctgctgagcgagcgagtgccg 1176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 CTGAGGCTCTCGCCGCGCTCTCTGCGGCGGCGCATCTCCGCTGAGGTGGGAGG 191

QY 1177 caggtgagcgagctcgagcgagtcgaatcccaagccttggttgtagctagctg 1236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 CCGGTGATGTGCCGTGGGCGCAGACAGATCCACCCTTAAGGATCGAGCGCCTG 251
```

RESULT 4  
LOCUS CNS01213 645 bp DNA GSS 26-JUL-1999  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN08C07 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL101589  
VERSION AL101589.1 GI:5613200  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Plasmid Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 645)  
Genoscope.

## REFERENCE

AUTHORS

TITLE

JOURNAL

## COMMENT

Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (BDGP) -  
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES  
source Location/Qualifiers  
1..645

```
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN08C07"
/note="end : T7"
BASE COUNT      28 a      26 c      85 g      92 t      414 others
ORIGIN
```

Query Match 3.1%; Score 43.8; DB 13; Length 645;  
Best Local Similarity 11.6%; Pred. No. 1.5;  
Matches 57; Conservative 123; Mismatches 311; Indels 0; Gaps 0;

```
QY 760 ttgtctggcggtattatagggtctactcgtgttcaaccgcgtctaacccgcacctggcga 819
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 TCKKCGGNNGNNNNNNNGGCGCANNNTGKKKKGGNGGNNGTGTGTGNNGGGGGG 157

QY 820 cgtgcatatccagctcgatctggtgagctagctgcggtgctctgtgtgagtgctt 879
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 GGNNNNNNGCANNNTTNTGCTTTNNNGCNNNNNSGCCNCCTCNCNNNTNTNSSSSNS 217

QY 880 ggaactgagccgttcgcgcgtgcgcgtgtgtgagcatcttggggcggtgaagacagcgc 939
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 NNNTNNNNNNNNNNNNNNNNSSGSSNNNNNGSSSSNNNSGSSSSSSNSNTSSN 277

QY 940 aacaacttcgagatcaatgctgctgctgctgctgcgcgagatgaatcaagatcgacctgg 999
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 NNSNTNSNNNNNNNNNNNNNNNNSSSTSTNTNSNTNSATGSSSSSSSSSSGCT 337

QY 1000 ctggagcatcgcccgagcaatccgaccttaacctggttatccctctatcaacagcg 1059
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 TBSGSSSSSSSSSSSSNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 397

QY 1060 cgcttacgagcttgagcgctactccgctccgctggaagtcgcggtgtgagggcgagcta 1119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 398 NTSTSTNSNSNSNTSSTSSSTNTSTSTSSSSNTSSSSNTSSSTTTNTSSASNNNSNN 457

QY 1120 ctggagcttgagcgcttgctgctgctgctgctgctgctgctgctgctgctgctgctg 1179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 458 NGNSSSSNAANSTSTNSNTTSSNNNSNCNTTSAANTSTNTTATGANNNNATSSSN 517

QY 1180 gtgagcgagctcgagcgagcgaatcccaagccttggtgatgctagctagctgacg 1239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 SNNSSNNNNNNNNNSGSGTSSGNTSSSCTGTBSSGNSNSNNNSCNGSNNCTGANSN 577

QY 1240 ggtttgagcag 1250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 SSSSTTTGSSSS 588
```

RESULT 5  
LOCUS CNS0073W 922 bp DNA GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR1AD09 of RPI-8 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL066784  
VERSION AL066784.1 GI:4945247  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 922)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT  
- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).





and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

# FEATURES

## source

1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPC1-98"  
/clone="BACR19D16"  
/note="end : TET3"  
BASE COUNT 120 a 61 c 172 t 511 others  
ORIGIN

Query Match 3.0%; Score 42.4; DB 13; Length 925;  
Best Local Similarity 14.3%; Pred. No. 3.7; Mismatches 145; Indels 4; Gaps 1;  
Matches 52; Conservative 163; Mismatches 145; Indels 4; Gaps 1;

QY 842 tgggtgactagtcgagtcgtctgtgtgctgtgactgtgacgctgtgctgtgc 901  
DB 912 BSCSSSSMSSTSSSSBSCSSSSBSSSSSTSSSSBSSSSSSSGTSSACVKN 853  
QY 902 ggcgtgtgacatcttctgtgtgagcgtgacagcaacaactcggatcatgtct 961  
DB 852 SSSCGCGCGGMABCMCMSSSSSCG---SASARKYKVRASGGAGKRGCGSGASASHSS 797  
QY 962 ggcgtgagtcgtgacagatcatcaagatcgtgctgtgtgacatcgcccggaata 1021  
DB 796 SBAACBSSSSSCASCMSSSSSSASSSSRSHSGGAGGASASSSSSSSASAGSVSS 737  
QY 1022 ccagcttaactgctgttaccctatcaacagcgcgttaccgctgtgagcgtct 1081  
DB 736 ASSSSSSSSSVSCSVASMSGSSSSSSASASSSSSSSACACSCCTTSMSCST 677  
QY 1082 aactcgctccgcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtt 1141  
DB 676 SASMSARSSSSSSSSSSSSASASSSSSSSSSSSSSSGACGASMSGGGSG 617  
QY 1142 ggcgtgctgctgacagcgtgacgctgtgagcgtgagcgtgagcgtgac 1201  
DB 616 SVASSGMSSSSVSSGGRSGSGGGGAGGSSSSSSGSGSVSCSGCMCRSC 557  
QY 1202 gcaaa 1205  
DB 556 SSAA 553  
RESULT 8  
CNS016BO/c 1201 bp DNA GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION BACN15M12 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL106542  
VERSION AL106542.1 GI:5622445  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Plasmid Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1201)  
GENOSCOPE.  
AUTHORS Direct Submission  
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRI cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
JOURNAL - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
- Web : <http://www.edgp.ebi.ac.uk> - This Drosophila Genome Project (EDGP) -  
collaboration with the European Drosophila Genome Project (EDGP) -  
library (Dros BAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton

and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

# FEATURES

## source

1..1201  
/organism="Drosophila melanogaster"  
/plasmid="pBelOBAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN15M12"  
/note="end : SP6"  
BASE COUNT 292 a 254 c 222 g 205 t 228 others  
ORIGIN

Query Match 3.0%; Score 42.4; DB 13; Length 1201;  
Best Local Similarity 29.4%; Pred. No. 4; Mismatches 136; Indels 0; Gaps 0;  
Matches 87; Conservative 73; Mismatches 136; Indels 0; Gaps 0;

QY 587 ccacctgtgctgtctatctgtgtccagacagcgtcctttctgacagcagtcgag 646  
DB 1186 CCCYCCCCCCCCCTCTTCTGCGCBYCBSSYCTCTTTTCTGCGCBSSGGGG 1127  
QY 647 tgltygcgctgtgtctcccaagctgtgagcgtgacagagtgagcagcttat 706  
DB 1126 KGCSSYSGSSGKKGCGCTKSKTKYGTCTTGKYGSGGGSGSGGMMVTSCCC 1067  
QY 707 gctcatctcaccacagtcgtgctgtgctgtgtgtgtgtgtgtgtgtgtgt 766  
DB 1066 CGCMCTTGBCCCSCCCGCKGKGKGGGGGCGGCTTHTSCGGGGGCTTKSKGKT 1007  
QY 767 ggcgttattaggtcttactgtttcaaccgcgtctaccacacgtgtgagcgtgagc 826  
DB 1006 GCKCTTTTTRKTKTGTTTCSSGTYTTCGVBSCSGSGSGSVWGGSCSCC 947  
QY 827 tattccagtcgtatttggtgacagtcgctgtgtgtgtgtgtgtgtgtgtgt 882  
DB 946 CETYBGHGTCTSKRTTTTCTTKKTYCTTGYCTTKYHRDCTYSM 891

## RESULT 9

CNS010C9 970 bp DNA GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION BACN03I05 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL098787  
VERSION AL098787.1 GI:5610398  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Plasmid Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 970)  
GENOSCOPE.  
AUTHORS Direct Submission  
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRI cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
JOURNAL - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
- Web : <http://www.edgp.ebi.ac.uk> - This Drosophila Genome Project (EDGP) -  
collaboration with the European Drosophila Genome Project (EDGP) -  
library (Dros BAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelOBAC11.

## FEATURES

### source

1..970  
/organism="Drosophila melanogaster"  
/plasmid="pBelOBAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"





Db 1158 CSKSGGCBGSKBKSBSYSSKSCSCCBGCGCCCCGSGSGSKSTCCSCYCHMYT 1099  
Qy 673 ctggcagatcctgcagcaagctgcagcaatctatctcagctacagctcagct 732  
Db 1098 KKGGGGKGGKBBKCTCYKKTBBGGGGKGGSTGTGTCTCCCCCCCCCCCCCCSSGGG 1039  
Qy 733 gctgcagctggttctgcagcaatcttctgcagcttctgcagcttctgcagct 785  
Db 1038 GSSSSSSGCGGCGCTMYCTTKCCCKTGGGCGCTKKGGGGGGGBMY 986  
RESULT 14  
Bg167536 1868 bp mRNA EST 06-FEB-2001  
LOCUS 602342745F1 NIH\_MGC\_89 Homo sapiens cDNA IMAGE:4452947 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg167536  
VERSION Bg167536.1 GI:12674165  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ARCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLNL0242 row: 1 column: 12.  
FEATURES  
source Location/Qualifiers  
1. 1868  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="4452947"  
/clone\_lib="NIH\_MGC\_89"  
/tissue\_type="hyponephroma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: PCMV-SPORE; Site: 1; NotI;  
Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC Library."  
BASE COUNT 599 a 624 c 370 g 274 t 1 others  
ORIGIN  
Query Match 3.0%; Score 41.8; DB 11; Length 1868;  
Best Local Similarity 48.5%; Pred. No. 6.4;  
Matches 115; Conservative 0; Mismatches 122; Indels 0; Gaps 0;  
Qy 644 ggggtgagcgagcgagcggcccaagctgagcgatcgctcagcaagctgagcaagt 703  
Db 1464 GGGTTGCGGCTGTGTGCTGCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1405  
Qy 704 tatgtctacccctacactacagctcgagctgagctgagctgagctgagctgagct 763  
Db 1404 CATTGATCATCTGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1345  
Qy 764 tctggagcttatctagagctcctcagctcagctcagctcagctcagctcagct 823  
Db 1344 GCTGTGATTAAGTGTGCGGGGATGATGATGATGATGATGATGATGATGATGAT 1285  
Qy 824 ggcctatcccaagctgagctgagctgagctgagctgagctgagctgagctgagct 880  
|||||

Db 1284 GCGTGTCCGACAGAGGAGGTGGTGTGTTGTGATCTAAGCTGCGTCCCATTTCTGG 1228  
RESULT 15  
CNS016KD 1100 bp DNA GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence Sp6 end of BAC  
DEFINITION BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL106855  
VERSION AL106855.1 GI:5624152  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Plasmid Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1100)  
AUTHORS Direct Submission  
JOURNAL Genoscope.  
COMMENT Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage.  
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.  
FEATURES  
source Location/Qualifiers  
1. 1100  
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db\_xref="taxon:7227"  
/clone\_lib="DrosBAC"  
/clone="BACN16D22"  
/note="end : Sp6"  
BASE COUNT 132 a 229 c 106 g 220 t 413 others  
ORIGIN  
Query Match 3.0%; Score 41.6; DB 13; Length 1100;  
Best Local Similarity 21.9%; Pred. No. 6.1;  
Matches 119; Conservative 182; Mismatches 241; Indels 2; Gaps 2;  
Qy 728 ggggtgagctgagctgagctgagctgagctgagctgagctgagctgagct 787  
Db 555 KCGCTGYCMGCTGKGKGTGGGCKGKSSTGGGSKGTGCTGCTGCTGCTGCTGCTG 614  
Qy 788 ggttcaaccccgctaccccgaccctgagcgagctgagctgagctgagctgagct 847  
Db 615 GTGTTTAKSKGKGGCCBCGSGSGCKSTCGSGSGGGGKSKSCSCCGGSGSGS 674  
Qy 848 gactagtcgagctgagctgagctgagctgagctgagctgagctgagctgagct 907  
Db 675 GCGSCSSSSCGKCGKSGKSCSCSBSGGYCCBGSSSKCCSGCGGCGCGCGGS 734  
Qy 908 ttagcatcttctgtgtgagcgagcaagcaagcaagcaagcaagcaagcaagcaag 966  
Db 735 CCGCGTCCGCGGBBGGGCGSYCGCGSKCCHCTKTCCTCCSGYKSKKCGCGTGTGC 794  
Qy 967 ggggtgagcgagctgagctgagctgagctgagctgagctgagctgagctgagct 1026  
Db 795 KCKCTGTGCGCBKCYCTTCKTSYCKBCKYKCGCCGCGSCBCKSCCKCTKCK 854  
Qy 1027 ttaacctggttaccctacacagcgagcgagctttagcgagctgagcgagct 1086  
Db 855 YKKKKKKYKBBKCKCKCTKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCK 914  
Qy 1087 gtcccgctgagagtcgagctgagcgagctgagctgagctgagctgagctgagct 1146  
|||||





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2002, 21:39:13 : Search time 47.17 Seconds  
(without alignments)  
6741.045 Million cell updates/sec

Title: US-09-887-038-2

Perfect score: 1404

Sequence: 1 atgactgtctgcaactctc.....attcagacgaaaaatgtag 1404

Scoring table:

IDEMTRY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PCrUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/backfiles1.seq:\*

\* Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.8	3.5	7218	1 US-08-232-463-14	Sequence 14, Appl
2	47.4	3.4	2232	1 US-08-241-766-12	Sequence 12, Appl
3	47.4	3.4	3120	1 US-08-491-146-1	Sequence 1, Appl
4	47.4	3.4	3120	1 US-08-241-766-11	Sequence 11, Appl
5	47.4	3.4	3120	1 US-08-234-011-1	Sequence 1, Appl
6	47.4	3.4	3120	2 US-08-701-062A-1	Sequence 1, Appl
7	39.6	2.8	1723	1 US-08-241-766-1	Sequence 1, Appl
8	39.6	2.8	1723	1 US-08-241-766-2	Sequence 2, Appl
9	35.8	2.5	2056	4 US-09-334-601-12	Sequence 12, Appl
10	35.2	2.5	390	4 US-09-911-853-20	Sequence 20, Appl
11	35.2	2.5	390	4 US-09-479-409-20	Sequence 20, Appl
12	35.2	2.5	710	4 US-08-998-416-706	Sequence 706, Appl
13	35.2	2.5	17612	3 US-08-911-853-29	Sequence 29, Appl
14	35.2	2.5	17612	4 US-09-479-409-29	Sequence 29, Appl
15	35	2.5	30001	1 US-08-125-468-1	Sequence 1, Appl
16	35	2.5	30001	2 US-08-474-933-1	Sequence 1, Appl
17	34.8	2.5	43280	2 US-08-804-227C-1	Sequence 1, Appl
18	34.4	2.5	4791	4 US-08-949-155-49	Sequence 49, Appl
19	33.6	2.4	33529	4 US-09-144-085-3	Sequence 3, Appl
20	33.4	2.4	4403765	4 US-09-103-840A-2	Sequence 2, Appl
21	33.4	2.4	4411529	4 US-09-103-840A-1	Sequence 1, Appl
22	33	2.4	326	1 US-08-700-575-3	Sequence 110, Appl
23	32.8	2.3	3410	4 US-09-020-956-110	Sequence 110, Appl
24	32.8	2.3	3410	4 US-09-030-607-110	Sequence 110, Appl
25	32.6	2.3	1088	4 US-09-077-675A-6	Sequence 6, Appl
26	32.6	2.3	1122	4 US-09-077-675A-9	Sequence 9, Appl
27	32.6	2.3	2221	4 US-09-232-191-24	Sequence 24, Appl

28	32.6	2.3	2221	4 US-09-232-200-74	Sequence 74, Appl
29	32.6	2.3	2221	4 US-09-232-197-74	Sequence 74, Appl
30	32.4	2.3	1272	3 US-09-036-987A-28	Sequence 28, Appl
31	32.4	2.3	1272	4 US-09-370-700-28	Sequence 28, Appl
32	32.4	2.3	80161	3 US-09-036-987A-1	Sequence 1, Appl
33	32.4	2.3	80161	4 US-09-370-700-1	Sequence 1, Appl
34	32.2	2.3	28958	1 US-08-258-618-6	Sequence 6, Appl
35	32.2	2.3	28958	1 US-08-456-837-6	Sequence 6, Appl
36	32.2	2.3	28958	1 US-08-457-342-6	Sequence 6, Appl
37	32.2	2.3	28958	1 US-08-457-646A-6	Sequence 6, Appl
38	32.2	2.3	28958	1 US-08-458-076A-6	Sequence 6, Appl
39	32.2	2.3	28958	1 US-08-764-233A-4	Sequence 4, Appl
40	32.2	2.3	28958	1 US-08-457-335A-6	Sequence 6, Appl
41	32.2	2.3	28958	1 US-08-729-214-6	Sequence 6, Appl
42	32.2	2.3	28958	3 US-09-028-934-6	Sequence 6, Appl
43	32.2	2.3	49377	1 US-08-764-233A-1	Sequence 1, Appl
44	32	2.3	6343	3 US-08-581-148C-30	Sequence 30, Appl
45	31.8	2.3	1548	2 US-08-762-106-5	Sequence 5, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOMLOPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
City: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PT99pt-Fls  
US-08-232-463-14





```
STRANDEDNESS: single
TOPOLOGY: both
MOLECULE TYPE: DNA
DESCRIPTION:
HYPOTHETICAL: NO
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: Inha operon
ORGANISM: M tuberculosis
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: M tuberculosis
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION: NO. 5556778e
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-491-146-1

Query Match          3.4%; Score 47.4; DB 1; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.00062;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagagcgtactcgtcccgctggaagtgcggttgaggcggaactactggcttga 1129
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1041 CGTCGATCGCGTTTCACATCGCACGCGGTAGCCGAGAGAGCGCGCCAGCTGCTCTCA 1100

QY 1130 cggccctcgcttgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1101 CCGGGTTCGACCGCGCTCGCGTGTATTCACGCCATCACCGCGCTCGCGCAAGGCC 1160

QY 1190 tgcggcgagatcgcaatcccaagcctttggtgatgctagcttggcggttgg 1246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1161 CGCTGCTCGACTCGACGTGCAAAAGAGAGACACCTGCGCAGCTTGCGCGCGGG 1217

RESULT 4
US-08-241-766-11
; Sequence 11, Application US/08241766
; Patent No. 5686590
; GENERAL INFORMATION:
; APPLICANT: JACOBS, W. R.
; APPLICANT: COLLINS, D. M.
; APPLICANT: BANERJEE, A.
; APPLICANT: DELISLE, G. M.
; APPLICANT: WILSON, T. M.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: AND TREATING MYCOBACTERIAL INFECTIONS USING AN Inha AGENT
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
```

```
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/241,766
FILING DATE: 12-May-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, GLADYS H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 25237-20003.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 3120 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-241-766-11

Query Match          3.4%; Score 47.4; DB 1; Length 3120;
Best Local Similarity 54.2%; Pred. No. 0.00062;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgttagagcgtactcgtcccgctggaagtgcggttgaggcggaactactggcttga 1129
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1041 CGTCGATCGCGTTTCACATCGCACGCGGTAGCCGAGAGAGCGCGCCAGCTGCTCTCA 1100

QY 1130 cggccctcgcttgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1101 CCGGGTTCGACCGCGCTCGCGTGTATTCACGCCATCACCGCGCTCGCGCAAGGCC 1160

QY 1190 tgcggcgagatcgcaatcccaagcctttggtgatgctagcttggcggttgg 1246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1161 CGCTGCTCGACTCGACGTGCAAAAGAGAGACACCTGCGCAGCTTGCGCGCGGG 1217

RESULT 5
US-08-234-011-1
; Sequence 1, Application US/08234011
; Patent No. 5702935
; GENERAL INFORMATION:
; APPLICANT: Sacchetti et al
; TITLE OF INVENTION: METHOD AND COMPOUNDS FOR INHIBITING
; TITLE OF INVENTION: LIPID BIOSYNTHESIS OF BACTERIA AND
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothenstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/234,011
FILING DATE: NO. 5702935 yet assigned
```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA: none  
ATTORNEY/AGENT INFORMATION:  
NAME: Pasqualini, Patricia A.  
REGISTRATION NUMBER: 34,894  
REFERENCE/DOCKET NUMBER: 96700/271  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3120  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: DNA  
DESCRIPTION: NO  
HYPOTHETICAL: NO  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: inhA operon  
ORGANISM: M tuberculosis  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE: M tuberculosis  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION: No. 5702935e  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-234-011-1

Query Match 3.4%; Score 47.4; DB 1; Length 3120;  
Best Local Similarity 54.2%; Pred. No. 0.00062;  
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1070 cgtttagcgccactccgctccgctggaaagtcgcggttgaaggcggaactactgggcttga 1129  
DB 1041 CGTCGATCGCGTTTCACATCGCAGCGGTAGCCAGGAGCGGCGCCCGCAGCTGCTCA 1100  
QY 1130 cggccttcgctggtgctgctgctgcaagcgcggtgacggcggtgctgcaagtgagccgac 1189  
DB 1101 CGGGGTTGACCGCGGCTGGCGGTGATTGACGGCATCAGCCGCGTCCGCGCAAGGCC 1160  
QY 1190 tgcggcgatcgcaatccccaagccttltgtgtgctagcttgccgggttgg 1246  
DB 1161 CGCTGCTGAACTCGACGTCGAAAGAGAGACACCTGCGCAGCTTGGCGCGCGG 1217

RESULT 6

US-08-701-062A-1  
Sequence 1, Application US/08701062A  
Patent No. 5882878  
GENERAL INFORMATION:  
APPLICANT: Sacchetti et al  
TITLE OF INVENTION: inhA CRYSTALS AND THREE  
TITLE OF INVENTION: DIMENSIONAL STRUCTURE  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESS: Amster, Rolfsstein & Edenstein  
STREET: 90 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Processor (ASCII)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,062A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: none  
ATTORNEY/AGENT INFORMATION:  
NAME: George, Kenneth P  
REGISTRATION NUMBER: 30,259  
REFERENCE/DOCKET NUMBER: 96700/296  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3120  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: both  
MOLECULE TYPE: DNA  
DESCRIPTION: NO  
HYPOTHETICAL: NO  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: inhA operon  
ORGANISM: M tuberculosis  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE: M tuberculosis  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION: No. 5882878e  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:





[illegible]

```

Db      205  CGGGGAGCCGCTGGACGCGAGCGCCGAGAACTGCGGGCTGCGGAACTCGACTTCTTG 264
QY      1150  ctgtctacggcggtgtgaacgagctgtggcaagctgaagccgactgcgcgcgatacgcaatccc 1209
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      265  CGGGTCTCTCGCGCGCGCGCTTGCGCGCGGCTGCACGCGACGAGCGCGGCGACTTGTGAGCC 324

RESULT 13
US-08-911-853-29
; Sequence 29, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Gerltse, Gijlsbert
; APPLICANT: Quax, Wilhelms J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,853
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17612 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-911-853-29

Query Match      2.5%; Score 35.2; DB 3; Length 17612;
Best Local Similarity 58.7%; Pred. No. 4.1;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY      1143  gctgctgtgtgtacagcgagtgacgagcggtgtgcgcaggtgagcagactgcgcgcgatacg 1202
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      14081  GATGCTGCGTGCACGACGAGGAGGTGCGGCTGCGCGAGCGCGGCTGCACAGGCAAGAT 14140

QY      1203  caatcccaagcctttgttgatgtgctagcttgcgcggtttgg 1246
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      14141  CGAAGACCATGCCCTGGTGACCTGATGAGTCCTTCGTCGGGCTGC 14184

RESULT 14
US-09-479-409-29
; Sequence 29, Application US/09479409
; Patent No. 6225106
; GENERAL INFORMATION:
; APPLICANT: Gerltse, Gijlsbert
; APPLICANT: Quax, Wilhelms J
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS

```



' . '

---

